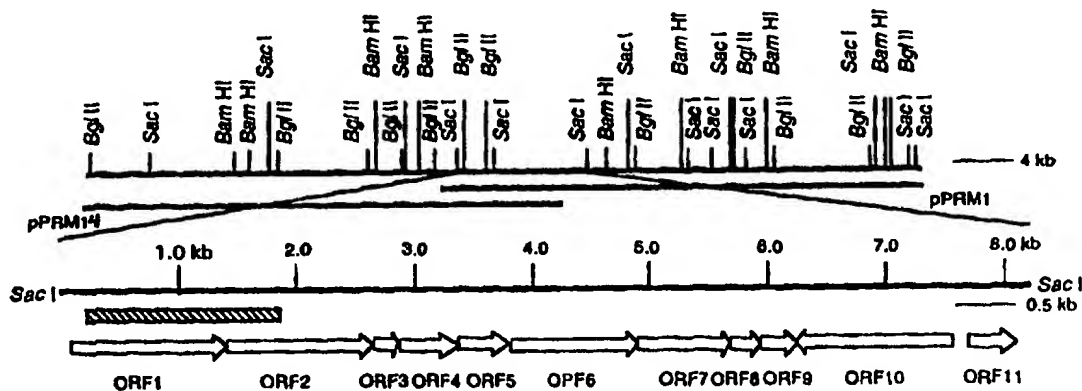




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : C12N 15/52, 1/21, 9/00, C12P 21/00 // (C12N 1/21, C12R 1:03)</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/11230</p> <p>(43) International Publication Date: 19 March 1998 (19.03.98)</p>
<p>(21) International Application Number: PCT/US96/14791</p> <p>(22) International Filing Date: 13 September 1996 (13.09.96)</p> <p>(71) Applicant: BRISTOL-MYERS SQUIBB COMPANY [US/US]; Route 206 - Province Line Road, P.O. Box 4000, Princeton, NJ 08543-4000 (US).</p> <p>(72) Inventors: OKI, Toshikazu; 4-20-10, Shodo, Sakea-ku, Yoko- hama 247 (JP). DAIRI, Tohru; 5-139, Tsukioka-cho, Toyama, Toyama 939 (JP).</p> <p>(74) Agent: BLOOM, Allen; Dechert Price & Rhoads, Princeton Pike Corporate Center, P.O. Box 5218, Princeton, NJ 08543- 5218 (US).</p>		<p>(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report.</i></p>

(54) Title: POLYKETIDE SYNTHASES FOR PRADIMICIN BIOSYNTHESIS AND DNA SEQUENCES ENCODING SAME



(57) Abstract

The present invention provides, *inter alia*, nucleic acids and corresponding amino acid sequences of several *Actinomadura* polyketide synthase genes that are useful, for example, in preparing pradimicin and analogs thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

POLYKETIDE SYNTHASES FOR PRADIMICIN BIOSYNTHESIS AND DNA SEQUENCES ENCODING SAME

The present invention relates, *inter alia*, to purified nucleic acids
5 encoding polyketide synthase genes for pradimicin biosynthesis, and
purified polypeptides having polyketide synthase activity. Polyketide
metabolites are natural products made by microorganisms and plants
from simple fatty acids. Many polyketides are used as human and
animal pharmaceuticals such as antibiotics, chemotherapeutics and
10 growth promoting agents, as well as flavoring agents and pigments.

Biosynthesis of polyketides is believed to occur by a series of
condensations of carbon units in a manner similar to that of long chain
fatty acids which are formed by fatty acid synthase. The fatty acids are
formed by a process in which a chain starter, usually a 2-carbon acetate
15 residue, which is joined by condensation to a chain extender unit, such
as malonate, to form an even-numbered chain. The resulting β -keto
group is then processed, by β -ketoacyl reduction, dehydration and enoyl
reduction. The cycle then begins again with the condensation of a new
extender unit. A typical fatty acid synthase is a multivalent system
20 involving eight functional units, acetyl, malonyl and palmityl
transferases, acyl carrier protein, ketoacyl synthase, ketoacyl reductase,
dehydratase and enoyl reductase. The organization of these units varies
in different organisms. See, for example, *EMBO J.* 8:2717-2725
(1989).

25 The fatty acid synthesis process differs from polyketide synthesis
since most polyketides contain structural complexities due to the use of
different starter and extender units, such as acetate, propionate and
butyrate. The polyketide synthesis is further complicated by variations
in the extent of processing of the β -carbon (β -ketoreduction,
30 dehydration, enoylreduction) as well as the introduction of chiral
carbons. See, for example, *Science* 252:675-679 (1991).

The tetracenomycin C polyketide synthase genes (*tmcl*) from
Streptomyces glaucescens, for example, have been sequenced, and the

sequence data revealed three complete open reading frames. An analysis of the sequence data resulted in a conclusion that polyketide synthesis in *S. glaucescens* involves a multienzyme complex consisting of at least five types of enzymes. These enzymes, which are
5 homologous to counterparts involved in fatty acid synthesis, are presumably involved in the assembly of the tetracenomycin C decaketide.

Additionally, for example, the structure and function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces*
10 *violaceoruber* has also been studied. This gene cluster has six open reading frames, thereby indicating that the granaticin-producing polyketide synthesis likely consists of at least six separate enzymes involved in carbon chain assembly. See *EMBO J.* 8:2717-2725 (1989). Further, *Streptomyces* polyketide synthase gene clusters involved in the
15 biosynthesis of actinorhodin and the *whiE* spore pigment have also been described. See *J. Biol. Chem.* 267:19278-19290 (1992) and *Gene* 130:107-116 (1993).

The molecular organization of the polyketide biosynthesis genes of *Saccharopolyspora erythraea*, which govern synthesis of the
20 polyketide portion of the macrolide antibiotic erythromycin, is similarly complex. The genes are organized in six repeated units that encode fatty acid synthase-like activities. Two repeated units are contained in a single open reading frame. It is believed that each repeated unit encodes a functional synthase unit and each synthase unit participates
25 in one of six fatty acid synthase-like elongation steps required for the formation of the polyketide. See *EMBO J.* 8:2727-2736 (1989).

Based on the above data, a model has been proposed in which polyketide genes have repeated units designated modules, and the corresponding proteins are called synthase units, wherein each synthase
30 unit is responsible for one of the fatty acid synthase-like cycles required for completing the polyketide. Thus, each synthase unit carries the

elements required for the condensation process, for selecting the particular extender unit to be incorporated, and for the extent of processing that the β -carbon will undergo. After completion of the cycle, the nascent polyketide is transferred from the acyl carrier protein (ACP) it occupies to the β -ketoacyl ACP synthase of the next synthase unit utilized, where the appropriate extender unit and processing level are introduced. This process is repeated, using a new synthase unit for each elongation cycle, until the programmed length has been reached. According to this model, formation of complex polyketides requires the participation of a different synthase unit for each cycle, thereby ensuring that the correct molecular structure is produced. See, for example, *Annu. Rev. Microbiol.* 47:875-912 (1993).

An actinomycete, namely, *Actinomadura*, certain strains of which were previously isolated from soil samples collected in the Fiji Islands and in India, was found to produce a complex of antibiotics designated pradimicin. See, for example, *J. Antibiot.* 43:755-762 (1990). Pradimicin A, as shown in Figure 1, has a unique dihydro-benzo[*a*]naphthacenequinone aglycon substituted with D-alanine and two sugars, and is a potent antifungal antibiotic produced, for example, by *Actinomadura hibisca* and *Actinomadura verrucosospora* subsp. *neohibisca*. See, for example, *J. Antibiot.* 43:755-762 (1990) and *J. Antibiot.* 46:387-397 (1993). Pradimicin is an antibiotic useful for multiple purposes, particularly for use as a pharmaceutical. For example, pradimicin has been shown to have activity against system fungal infections caused by *Candida albicans*, *Aspergillus fumigatus* and *Cryptococcus neoformans*. Further, pradimicin is active *in vitro* against a wide variety of fungi and yeasts, some Gram-positive bacteria, and viruses. *J. Org. Chem.* 54:2536-2539 (1989). Purified polypeptides having polyketide synthase activity and purified nucleic acids encoding such polypeptides are therefore desirable, for example, to provide pharmaceutically useful products.

SUMMARY OF THE INVENTION

Until now, the sequences encoding polyketide synthase genes in *Actinomadura* had not been identified. These sequences are provided in the present invention.

- 5 One preferred embodiment of the present invention is a substantially pure nucleic acid comprising a nucleic acid sharing at least about 75% nucleic acid identity with an open reading frame (ORF) of an *Actinomadura* polyketide synthase gene, and more preferably, at least about 80% identity, and most preferably, at least about 90% identity.
- 10 In certain preferred embodiments, the nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NO:1-12. A further preferred embodiment is a substantially pure nucleic acid comprising a nucleic acid encoding an *Actinomadura* polyketide synthase gene sharing at least about 75% amino acid identity, and more preferably, at
- 15 least about 80% identity, and most preferably, at least about 90% identity with a polypeptide encoded by a nucleic acid selected from the group consisting of SEQ ID NO:1-12.

- In certain preferred embodiments, the substantially pure nucleic acid comprises a nucleic acid encoding a polypeptide differing from an
- 20 *Actinomadura* polyketide synthase gene by no more than about 20 amino acid substitutions, and more preferably, no more than about 10 amino acid substitutions. Preferably, the substitutions cause a conservative substitution in the amino acid sequence of the encoded polyketide synthase. The nucleic acids of the invention also include
- 25 nucleic acid analogs.

- Further, the present invention provides a substantially pure nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 75% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone. Preferably, the nucleic
- 30 acid encodes a polypeptide sharing at least about 80%, and more preferably, at least about 90% amino acid identity with a polyketide

synthase for biosynthesis of a benzo(g)naphthacenequinone. In preferred embodiments, the polyketide synthase is an *Actinomadura* polyketide synthase, and the polyketide is preferably a dihydrobenzo(g)naphthacenequinone aglycon, and preferably pradimicin, such as Pradimicin A, B, C, D, E, FA-1, FA-2, FL, FS, H, 11-O-L-xylosylpradimicin H, L, S, T1, T2 or BMS181184.

Yet another embodiment of the invention is a substantially pure nucleic acid comprising a nucleic acid that hybridizes, under stringent conditions, to a nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 75% amino acid identity with an actinomadura polyketide synthase. More preferably, the nucleic acid hybridizes to a nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 80% amino acid identity with an *Actinomadura* polyketide synthase, and even more preferably, encoding a polypeptide sharing at least about 90% amino acid identity with an *Actinomadura* polyketide synthase. Most preferably, the nucleic acid hybridizes with a nucleic acid comprising a nucleic acid selected from the group consisting of SEQ ID NO:1-12. Such a hybridizing nucleic acid can be used, for example, to screen for organisms that produce pradimicin.

The invention additionally includes vectors capable of reproducing in a eukaryotic or prokaryotic cell having a nucleic acid described above as well as transformed eukaryotic or prokaryotic cells having such nucleic acid.

Thus, another preferred embodiment is a transformed eukaryotic or prokaryotic cell comprising a nucleic acid encoding a polypeptide sharing at least about 70% amino acid identity with an *Actinomadura* polyketide synthase gene, and more preferably, at least about 80% identity, and most preferably, at least about 90% identity. Most preferably, the nucleic acid sequence comprises a nucleic acid selected from the group consisting of SEQ ID NO:1-12. Preferably, the

transformed cell expresses one of the *Actinomadura* polyketide synthase genes described herein.

Yet another preferred embodiment is a vector capable of reproducing in a eukaryotic or prokaryotic cell comprising a nucleic acid
5 encoding a polypeptide sharing at least about 70% nucleic acid identity with an *Actinomadura* polyketide synthase gene, and more preferably, at least about 80% identity, and most preferably, at least about 90% identity. Preferably, the nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NO:1-12. Preferably, the inventive
10 vector expresses, intracellularly or extracellularly, one of the *Actinomadura* polyketide synthases described herein.

Another embodiment of the present invention provides a substantially pure polypeptide comprising an amino acid sequence sharing at least about 75% amino acid identity with an *Actinomadura*
15 polyketide synthase, and more preferably, at least about 80% identity, and most preferably, at least about 90% identity. Preferably, the polypeptide shares at least about 75% amino acid identity with a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:13-15.

20 Yet another preferred embodiment is a method of preparing pradimicin or a pradimicin analog thereof, comprising transforming a eukaryotic or prokaryotic cell with an expression vector for expressing intracellularly or extracellularly a nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 70% amino acid identity
25 with an *Actinomadura* polyketide synthase, growing the transformed cell in culture, and isolating the pradimicin or analog thereof from the transformed cell or the culture medium. Preferably, the polypeptide shares at least about 80% amino acid identity with an *Actinomadura* polyketide synthase, and more preferably, the polypeptide shares at
30 least about 90% amino acid identity with an *Actinomadura* polyketide synthase. Most preferably, the expression vector comprises a nucleic

acid encoding all polyketide synthase genes necessary for synthesis of pradimicin, such as SEQ ID NO:1.

BRIEF DESCRIPTION OF THE DRAWINGS

5 **Figure 1** shows the chemical structure of two types of pradimicin, pradimicin A and pradimicin S.

Figure 2 shows conserved amino acid sequences in β -ketosynthases and acyl transferases for granaticin, tetracenomycin and
10 actinorhodin. These conserved sequences were used to create two probes for cloning the polyketide synthase genes in *Actinomadura*.

Figure 3 shows a restriction map of *Actinomadura* polyketide synthase genes, ORFs 1-11.

15

Figure 4 provides an alignment of the *Actinomadura* ORF1 gene product ("A") (SEQ ID NO:13) with a *Streptomyces* polyketide synthase gene product for tetracenomycin biosynthesis ("B").

20 **Figure 5** provides an alignment of the *Actinomadura* ORF2 gene product ("A") (SEQ ID NO:14) with a *Streptomyces* polyketide synthase gene product for actinorhodin biosynthesis ("B").

DETAILED DESCRIPTION

25

 The present invention provides, *inter alia*, nucleic acids and corresponding amino acid sequences of *Actinomadura* polyketide synthase genes. The polyketide synthases are responsible for the biosynthesis of pradimicin, such as zwitterionic pradimicins A, B and C,
30 which are produced, for example, by *Actinomadura hibisca*, and pradimicin S, which is produced, for example, by *Actinomadura spinosa*.

See Figure 1, which provides the chemical structures of pradimicins A and S. See also *J. Antibiot.* 43:755-762 (1990). Pradimicin is useful, for example, as an antibiotic, including use as an anti-fungal and an anti-viral agent. For example, pradimicin has been shown to have activity
5 against system fungal infections caused by *Candida albicans*, *Aspergillus fumigatus* and *Cryptococcus neoformans*. Further, pradimicin is active *in vitro* against a wide variety of fungi and yeasts, some Gram-positive bacteria, and viruses. *J. Org. Chem.* 54:2536-2539 (1989). For instance, pradimicin is believed to be active against HIV. See, for
10 example, *J. Antibiot.* 41:1708 (1988) and *Virology* 176:467 (1990).

Techniques used in the prior art were not applicable for cloning pradimicin A biosynthetic genes from *Actinomadura hibisca*. Specifically, many antibiotic biosynthetic genes including self-defense genes in actinomycetes are clustered in a genomic region. The close
15 linkage between antibiotic biosynthetic genes and self-defense genes has provided a useful tool for cloning of antibiotic biosynthetic genes, since transformants carrying antibiotic resistance determinants can be selected. However, this technique could not be applied to the cloning of the pradimicin A biosynthetic gene cluster because pradimicin A had not
20 been shown to have significant antibacterial activity. Therefore, the polyketide synthase genes for pradimicin A biosynthesis were cloned from *Actinomadura hibisca* using oligonucleotide probes based on the conserved amino acid sequences of other polyketide synthase genes, followed by cloning of the flanking region of pradimicin A polyketide
25 synthase genes. Specifically, certain amino acid sequences of β -keto synthase, acyl transferase and acyl carrier protein of polyketide synthases are strongly conserved in *Streptomyces* strains producing polyketide antibiotics. See *Annu. Rev. Microbiol.* 47:875-912 (1993) and *J. Biol. Chem.* 267:19278-19290 (1992). Based on these
30 sequences, two oligonucleotide probes were synthesized, as shown in

Figure 2. See also Example 1, which provides experimental details of the cloning of the pradimicin A polyketide synthase genes.

After screening with an *Actinomadura hibisca* library, an 8.2 kb *Sac* I fragment was identified, which hybridized with these
 5 oligonucleotide probes. By DNA sequencing of the 8.2 kb *Sac* I fragment (SEQ ID NO:1), eleven open reading frames (ORFs) were identified. All of ORFs except for ORF10 are believed to be translated in the same direction. Referring to SEQ ID NO:1, ORF1 spans from position 72 (beginning with GTG) to position 1347 (ending with TGA);
 10 ORF2 spans from 1346 (GTG) to 2567 (TGA); ORF3 spans from 2594 (ATG) to 2855 (TGA); ORF4 spans from 2854 (ATG) to 3313 (TGA); ORF5 spans from 3312 (GTG) to 3771 (TGA); ORF6 spans from 3794 (ATG) to 4817 (TGA); ORF7 spans from 4857 (ATG) to 5595 (TGA); ORF8 spans from 5594 (GTG) to 5933 (TGA); ORF9 spans from 5932
 15 (GTG) to 6241 (TAA); ORF10 spans, in reverse direction, from 7534 (ATG) to 6301 (TAG) and ORF11 spans from 7668 (ATG) to 8010 (TGA).

Each of the deduced ORFs has a significant similarity to a protein responsible for polyketide biosynthesis or spore color formation in other
 20 organisms. ORF1, ORF2 and ORF3 have particularly strong similarities (50% - 70% amino acid identity) with polyketide synthases for actinorhodin biosynthesis. See, for example, Figure 4, which provides an alignment of the ORF1 gene product with a *Streptomyces* polyketide synthase gene product for tetracenomycin biosynthesis, and Figure 5,
 25 which provides an alignment of the ORF2 gene product with a *Streptomyces* polyketide synthase gene product for actinorhodin biosynthesis. See also Table 1 below.

Table 1

30	ORFs	Number of	Molecular	Translational	Homologous proteins
		amino acids	weight	coupling	

	ORF1	426	44,440	Unknown	Hypothetical protein 4 of <i>Sac. hirsuta</i> (73% identity among 413 amino acids) ¹¹ <i>tcm la</i> gene of <i>S. glaucescens</i> (73%/412) ²¹ <i>gra I</i> gene of <i>S. violacearuber</i> (71%/413) ³¹ <i>sct I</i> ORF1 of <i>S. coelicolor</i> (69%/415) ⁴¹
	ORF2	408	41,610	ORF1/ORF2	<i>sct I</i> ORF2 of <i>S. coelicolor</i> (57%/397) ⁴¹ <i>tcm Id</i> gene of <i>S. glaucescens</i> (54%/403) ²¹ Beta-ketoacyl synthase chain 2 of <i>S. cinnamonensis</i> (50%/397) ⁵¹
	ORF3	88	9,888	—	Hypothetical protein 6 of <i>Sac. hirsuta</i> (51%/78) ¹¹ Granaticin-producing PKS acyl carrier protein of <i>S. violacearuber</i> (53%/75) ³¹ Actinorhodin-producing PKS acyl carrier protein of <i>S. coelicolor</i> (51%/75) ⁴¹
	ORF4	154	17,694	ORF3/ORF4	Hypothetical protein 7 of <i>S. coelicolor</i> (58%/149) ⁶¹ PKS cyclase <i>curF</i> of <i>S. cyaneus</i> (61%/142) ⁷¹ <i>tcmN</i> protein of <i>S. glaucescens</i> (52%/149) ⁸¹
5	ORF5	154	15,784	ORF4/ORF5	Hypothetical protein 6 of <i>Mixococcus xanthus</i> (46%/39) ⁸¹ Histidine protein kinase <i>divJ</i> of <i>Caulobacter crescentus</i> (26%/102) ¹⁰¹ Multicatalytic endopeptidase complex chain Y7 of <i>Sac. cerevisiae</i> (23%/105) ¹¹¹
	ORF6	342	37,004	—	<i>tcmN</i> protein of <i>S. glaucescens</i> (47%/330) ⁸¹ Carminomycin 4-O-methyltransferase of <i>S. peucetius</i> (30%/317) ¹²¹ O-demethylpuromycin O-methyltransferase of <i>S. anulatus</i> (33%/334) ¹³¹
	ORF7	247	25,583	—	3-ketoacyl-ACP reductase <i>fab G</i> of <i>E. coli</i> (38%/244) ¹⁴¹ Granaticin-producing PKS chain 5 of <i>S. violacearuber</i> (30%/251) ³¹ Granaticin-producing PKS chain 6 of <i>S. violacearuber</i> (35%/252) ³¹
	ORF8	114	12,986	ORF7/ORF8	Hypothetical protein 1 of <i>S. coelicolor</i> (24%/80) ⁶¹

ORF9	104	11,279	ORF8/ORF9	Hypothetical protein 1 of <i>S. coelicolor</i> (24%/91) ⁶¹ Hypothetical protein 6 of <i>Sac. hirsuta</i> (27%/48) ¹¹ Hypothetical 41.2 KD protein of <i>S. halstedii</i> (24%/91) ¹⁶¹
ORF10	412	44,857	—	Cytochrome P450 105B1 of <i>S. griseolus</i> (40%/404) ¹⁶¹ Cytochrome P450 P450CVIIB1 of <i>Sac. erythraea</i> (38%/405) ¹⁷¹ Cytochrome P450 105C1 of <i>Streptomyces</i> sp. (41%/323) ¹⁸¹
ORF11	115	13,036	—	Hypothetical protein 7 of <i>S. coelicolor</i> (51% 107) ⁶¹ <i>curG</i> protein of <i>S. cyaneus</i> (45%/106) ⁷¹ <i>tcml</i> protein of <i>S. glaucescens</i> (35%/105) ¹⁹¹

5

¹¹ *Mol. Gen. Genet.* 240:146-150 (1993).

²¹ *EMBO J.* 8:2727-2736 (1989).

³¹ *EMBO J.* 8:2717-2725 (1989).

⁴¹ *J. Biol. Chem.* 267:19278-19290 (1992).

10 ⁵¹ *Mol. Gen. Genet.* 234:254-264 (1992).

⁶¹ *Mol. Microbiol.* 4:1679-1691 (1990).

⁷¹ *Gene* 117:131-136 (1992).

⁸¹ *J. Bacteriol.* 174:1810-1820 (1992).

⁹¹ EMBL data library no. S32173.

15 ¹⁰¹ *Proc. Natl. Acad. Sci.* 89:10297-10301 (1992).

¹¹¹ *Mol. Cell. Biol.* 11:344-353 (1991).

¹²¹ *J. Bacteriol.* 175:3900-3904 (1993).

¹³¹ *Gene* 109:55-61 (1991).

¹⁴¹ *J. Biol. Chem.* 267:5751-5754 (1992).

20 ¹⁵¹ *Gene* 130:107-116 (1993).

¹⁶¹ *J. Bacteriol.* 173:3335-3345 (1990).

¹⁷¹ *J. Bacteriol.* 174:725-735 (1992).

¹⁸¹ *J. Bacteriol.* 172:3644-3653 (1990).

¹⁹⁾ EMBL data library no. S27691.

DNA regions homologous to the *Actinomadura* polyketide synthase genes were specifically found in all of pradimicin producers
5 examined, but not in pradimicin non-producers in genomic Southern hybridization, thereby providing evidence that the genes cloned encode polyketide synthases for pradimicin biosynthesis.

Thus, the present invention provides, *inter alia*, nucleic acids encoding *Actinomadura* polyketide synthase genes and polypeptides and
10 analogs thereof, including nucleic acids that bind to an *Actinomadura* polyketide synthase gene. The nucleic acids can be used, for example, to screen for organisms that produce pradimicin or that have homologous polyketide synthase gene sequences. Further, the nucleic acids can be used, for instance, to synthesize polyketide synthases,
15 which can in turn be used, for example, to produce pradimicin.

The *Actinomadura* species include but are not limited to *Actinomadura hibisca*, *Actinomadura verrucosospora*, and particularly subsp. *neohibisca*, *Actinomadura libanotica*, *Actinomadura echinospora*, *Actinomadura chengduensis*, *Actinomadura kijaniata*, *Actinomadura*
20 *atramentaria*, *Actinomadura citrea*, *Actinomadura cremea*, *Actinomadura fulvescens*, *Actinomadura viridis*, *Actinomadura roseoviolacea*, *Actinomadura verrucosopora*, *Actinomadura madurae*, *Actinomadura pelletieri* and, for example, other soil isolates.

25 1. Nucleic Acids

The present invention provides, *inter alia*, nucleic acids. The nucleic acid embodiments of the invention are preferably deoxyribonucleic acids (DNAs), both single- and double-stranded, and most preferably double-stranded deoxyribonucleic acids. However, they
30 can also be ribonucleic acids (RNAs), as well as hybrid RNA:DNA double-stranded molecules.

Nucleic acids encoding an *Actinomadura* polyketide synthase gene include all *Actinomadura* polyketide synthase gene-encoding nucleic acids, whether native or synthetic, RNA, DNA, or cDNA, that encode an *Actinomadura* polyketide synthase gene, or the complementary strand thereof, including but not limited to nucleic acid found in an *Actinomadura* polyketide synthase gene-expressing organism. For recombinant expression purposes, codon usage preferences for the organism in which such a nucleic acid is to be expressed are advantageously considered in designing a synthetic polyketide synthase-encoding nucleic acid.

Further, the present invention provides a substantially pure nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 75% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone. Preferably, the nucleic acid encodes a polypeptide sharing at least about 80%, and more preferably, at least about 90% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone. In preferred embodiments, the polyketide synthase is an *Actinomadura* polyketide synthase, and the polyketide is preferably a dihydrobenzo(a)naphthacenequinone aglycon, and preferably pradimicin, such as Pradimicin A, B, C, D, E, FA-1, FA-2, FL, FS, H, 11-O-L-xylosylpradimicin H, L, S, T1, T2 or BMS181184. For a description of the foregoing pradimicins, see, for example, *J. Antibiot.* 41:1701 (1988), *J. Org. Chem.* 54:2536 (1989), *J. Antibiot.* 43:771 (1990), *J. Antibiot.* 43:1223 (1990), *J. Antibiot.* 46:265 (1993), *J. Antibiot.* 46:398 (1993), *J. Antibiot.* 46:406 (1993), *J. Antibiot.* 46:598 (1993), and *J. Antibiot.* 46:1589 (1993).

In addition to nucleic acids encoding an *Actinomadura* polyketide synthase gene, the present invention includes nucleic acids encoding polypeptides that are homologous to or share a percentage amino acid identity with *Actinomadura* polyketide synthases.

Numerous methods for determining percent homology are known in the art. One preferred method is to use version 6.0 of the GAP computer program for making sequence comparisons. The program is available from the University of Wisconsin Genetics Computer Group and utilizes the alignment method of Needleman and Wunsch, *J. Mol. Biol.* 48, 443, 1970, as revised by Smith and Waterman *Adv. Appl. Math.* 2, 482, 1981.

Numerous methods for determining percent identity are also known in the art, such as use of the FASTA computer program, which is also available from the University of Wisconsin. Preferably, the program used to determine percent identity is the DNASIS program, which is available from Hitachi Corp. (Tokyo, Japan).

To construct non-naturally occurring *Actinomadura* polyketide synthase gene-encoding nucleic acids, the native sequences can be used as a starting point and modified to suit particular needs. The nucleic acids of the invention include, for example, the nucleic acids of SEQ ID NO:1-12.

The invention is also directed to a nucleic acid encoding a segment of an *Actinomadura* polyketide synthase gene. Preferably, the encoded polypeptide will be effective to perform its function, such as an enzymatic function, that is performed by the full-size polyketide synthase.

For identifying the active domain or domains of *Actinomadura* polyketide synthase genes, one approach is to take an *Actinomadura* polyketide synthase gene cDNA and create deletional mutants lacking segments at either the 5' or the 3' end by, for instance, partial digestion with S1 nuclease, Bal 31 or Mung Bean nuclease (the latter approach described in literature available from Stratagene, San Diego, CA, in connection with a commercial deletion cloning kit). Alternatively, the deletion mutants are constructed by subcloning restriction fragments of an *Actinomadura* polyketide synthase gene cDNA. The deletional

constructs are cloned into expression vectors and tested for their polyketide synthase activity.

These structural genes can be altered by mutagenesis methods such as that described by Adelman et al., *DNA*, 2: 183 (1983) or
5 through the use of synthetic nucleic acid strands. The products of mutant genes can be tested for polyketide synthase activity.

The nucleic acid sequences can be further mutated, for example, to incorporate useful restriction sites. See Maniatis et al. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Press, 1989). Such
10 restriction sites can be used to create "cassettes," or regions of nucleic acid sequence that are facilely substituted using restriction enzymes and ligation reactions. The cassettes can be used to substitute synthetic sequences encoding mutated *Actinomadura* polyketide synthase amino acid sequences.

15 *Actinomadura* polyketide synthase gene-encoding sequences can be, for instance, substantially or fully synthetic. See, for example, Goeddel et al., *Proc. Natl. Acad. Sci. USA*, 76, 106-110 (1979). For recombinant expression purposes, codon usage preferences for the organism in which such a nucleic acid is to be expressed are
20 advantageously considered in designing a synthetic *Actinomadura* polyketide synthase gene-encoding nucleic acid. Since the nucleic acid code is degenerate, numerous nucleic acid sequences can be used to create the same amino acid sequence.

Further, with an altered amino acid sequence, numerous methods
25 are known to delete sequences from or mutate nucleic acid sequences that encode a polypeptide and to confirm the function of the polypeptides encoded by these deleted or mutated sequences. Accordingly, the invention also relates to a mutated or deleted version of an *Actinomadura* polyketide synthase nucleic acid that encodes a
30 polypeptide that preferably retains polyketide synthase activity.

Conservative mutations are preferred. Such conservative mutations include mutations that switch one amino acid for another within one of the following groups:

1. Small aliphatic, nonpolar or slightly polar residues: Ala, Ser, Thr,
5 Pro and Gly;
2. Polar, negatively charged residues and their amides: Asp, Asn,
Glu and Gln;
3. Polar, positively charged residues: His, Arg and Lys;
4. Large aliphatic, nonpolar residues: Met, Leu, Ile, Val and Cys;
10 and
5. Aromatic residues: Phe, Tyr and Trp.

A preferred listing of conservative substitutions is the following:

15	Original Residue	Substitution
	Ala	Gly, Ser
	Arg	Lys
	Asn	Gln, His
	Asp	Glu
20	Cys	Ser
	Gln	Asn
	Glu	Asp
	Gly	Ala, Pro
	His	Asn, Gln
25	Ile	Leu, Val

	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Tyr, Ile
	Phe	Met, Leu, Tyr
5	Ser	Thr
	Thr	Ser
	Trp	Tyr
	Tyr	Trp, Phe
10	Val	Ile, Leu

The types of substitutions selected may be based on the analysis of the frequencies of amino acid substitutions between homologous proteins of different species developed by Schulz et al., *Principles of Protein*
 15 *Structure*, (Springer-Verlag, 1978), pp. 14-16, on the analyses of structure-forming potentials developed by Chou and Fasman, *Biochemistry* 13: 211 (1974) or other such methods reviewed by Schulz et al, *Principles in Protein Structure*, (Springer-Verlag, 1978), pp. 108-130, and on the analysis of hydrophobicity patterns in proteins
 20 developed by Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132 (1982).

2. Polypeptides

In addition to analogs of nucleic acid sequences, the present invention includes analogs of *Actinomadura* polyketide synthases that
 25 preferably retain polyketide synthase activity. Preferably, the analogs will share at least about 75% amino acid identity, more preferably, at least about 80% identity, even more preferably, at least about 85%

identity, even more preferably at least about 90% identity, and most preferably at least about 95% identity to an *Actinomadura* polyketide synthase, such as the polypeptide of SEQ ID NO:13, SEQ ID NO:14 or SEQ ID NO:15.

5

3. Methods of Synthesizing Polypeptides

In one embodiment, the polypeptides of the invention are made as follows, using a gene fusion. For example, fusion to maltose-binding protein ("MBP") can be used to facilitate the expression and purification
10 of a polyketide synthase in a prokaryote such as *E.coli*. The hybrid protein can be purified, for example, using affinity chromatography using the binding protein's substrate. See, for example, *Gene* 67: 21-30 (1988). When using a fusion protein that includes maltose binding protein, a cross-linked amylose affinity chromatography column can be
15 used to purify the protein.

The cDNA specific for a given polyketide synthase or analog thereof can also be linked using standard means to a cDNA for glutathione S-transferase ("GST"), found on a commercial vector, for example. The fusion protein expressed by such a vector construct
20 includes the polyketide synthase or analog and GST, and can be treated for purification.

Should the MBP or GST portion of the fusion protein interfere with function, it is removed by partial proteolytic digestion approaches that preferentially attack unstructured regions, such as the linkers
25 between MBP or GST and the polyketide synthase. The linkers are designed to lack structure, for instance using the rules for secondary structure-forming potential developed by Chou and Fasman, *Biochemistry* 13, 211, 1974. The linker is also designed to incorporate protease target amino acids, such as trypsin, arginine and lysine
30 residues. To create the linkers, standard synthetic approaches for making oligonucleotides are employed together with standard subcloning

methodologies. Other fusion partners other than GST or MBP can also be used.

Additionally, the *Actinomadura* polyketide synthases can be directly synthesized from nucleic acid (by the cellular machinery)

- 5 without use of fusion partners. For instance, nucleic acids having the sequence of any of SEQ ID NO:1-12 are subcloned into an appropriate expression vector having an appropriate promoter and expressed in an appropriate organism. Antibodies against *Actinomadura* polyketide synthases can be employed to facilitate purification.

- 10 Additional purifications techniques are applied as needed, including without limitation, preparative electrophoresis, FPLC (Pharmacia, Uppsala, Sweden), HPLC (e.g., using gel filtration, reverse-phase or mildly hydrophobic columns), gel filtration, differential precipitation (for instance, "salting out" precipitations), ion-exchange
15 chromatography and affinity chromatography (including affinity chromatography using the RE1 duplex nucleotide sequence as the affinity ligand).

- A polypeptide or nucleic acid is "isolated" in accordance with the invention in that the molecular cloning of the nucleic acid of interest, for
20 example, involves taking an *Actinomadura* polyketide synthase gene nucleic acid from a cell, and isolating it from other nucleic acids. This isolated nucleic acid may then be inserted into a host cell, which may be yeast or bacteria, for example. A polypeptide or nucleic acid is "substantially pure" in accordance with the invention if it is
25 predominantly free of other polypeptides or nucleic acids, respectively. A macromolecule, such as a nucleic acid or a polypeptide, is predominantly free of other polypeptides or nucleic acids if it constitutes at least about 50% by weight of the given macromolecule in a composition. Preferably, the polypeptide or nucleic acid of the present
30 invention constitutes at least about 60% by weight of the total polypeptides or nucleic acids, respectively, that are present in a given

composition thereof, more preferably about 80%, still more preferably about 90%, yet more preferably about 95%, and most preferably about 100%. Such compositions are referred to herein as being polypeptides or nucleic acids that are 60% pure, 80% pure, 90% pure, 95% pure, or
5 100% pure, any of which are substantially pure.

4. Means for Identifying Polypeptides with *Actinomadura* Polyketide Synthase Activity

In one aspect, the present invention provides methods for
10 identifying polypeptides that are homologous to an *Actinomadura* polyketide synthase using an *Actinomadura* polyketide synthase cDNA, for example.

Additionally, probes for *Actinomadura* polyketide synthase expression can be used, for example, to detect the presence of an
15 *Actinomadura* polyketide synthase. Such probes include antibodies directed against an *Actinomadura* polyketide synthase or fragments thereof, nucleic acid probes that hybridize, under stringent conditions, to an *Actinomadura* polyketide synthase mRNA, and oligonucleotides that specifically prime a PCR amplification of an *Actinomadura* polyketide
20 synthase mRNA. Nucleic acid molecules that bind to an *Actinomadura* polyketide-encoding nucleic acid under high stringency conditions are identified functionally, or by using the hybridization rules reviewed in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Press, 1989).

25 Many deletional or mutational analogs of nucleic acid sequences for an *Actinomadura* polyketide synthase are effective hybridization probes for *Actinomadura* polyketide synthase-encoding nucleic acid. Accordingly, the present invention relates to nucleic acids that hybridize with such *Actinomadura* polyketide synthase-encoding nucleic acids
30 under stringent conditions. Preferably, the nucleic acid of the present

invention hybridizes, under stringent conditions, with at least a segment of any of the nucleic acids described as SEQ ID NO:1-12.

"Stringent conditions" refers to conditions that allow for the hybridization of substantially related nucleic acids, where relatedness is a function of the sequence of nucleotides in the respective nucleic acids. For instance, for a nucleic acid of 100 nucleotides, such conditions will generally allow hybridization thereto of a second nucleic acid having at least about 85% homology, and more preferably having at least about 90% homology. Such hybridization conditions are described by Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Press, 1989).

PCR (polymerase chain reaction) can be used to detect nucleic acids having *Actinomadura* polyketide synthase sequences through amplification of such sequences using *Actinomadura* polyketide synthase nucleic acid primers. PCR methods of amplifying nucleic acids utilize at least two primers. One of these primers is capable of hybridizing to a first strand of the nucleic acid to be amplified and of priming enzyme-driven nucleic acid synthesis in a first direction. The other is capable of hybridizing the reciprocal sequence of the first strand (if the sequence to be amplified is single stranded, this sequence is initially hypothetical, but is synthesized in the first amplification cycle) and of priming nucleic acid synthesis from that strand in the direction opposite the first direction and towards the site of hybridization for the first primer. Conditions for conducting such amplifications, particularly under preferred high stringency conditions, are well known. See, for example, *PCR Protocols* (Cold Spring Harbor Press, 1991).

Antibodies against *Actinomadura* polyketide synthases can also be used to identify polypeptides that are homologous to *Actinomadura* polyketide synthases. Antigens for eliciting the production of antibodies against an *Actinomadura* polyketide synthase can be produced recombinantly by expressing all of or a part of the nucleic acid of an

Actinomadura polyketide synthase in a bacteria or a yeast or other eukaryotic cell line. In one embodiment, the recombinant protein is expressed as a fusion protein, with the non-*Actinomadura* polyketide synthase portion of the protein serving either to facilitate purification or to enhance the immunogenicity of the fusion protein. For instance, the non-*Actinomadura* polyketide synthase portion comprises a protein for which there is a readily-available binding partner that is utilized for affinity purification of the fusion protein. The antigen includes an "antigenic determinant," i.e., a minimum portion of amino acids sufficient to bind specifically with an anti-*Actinomadura* polyketide synthase antibody.

Antisera to an *Actinomadura* polyketide synthase can be made, for example, by creating an *Actinomadura* polyketide synthase antigen by linking a portion of the cDNA for *Actinomadura* polyketide synthase to a cDNA for glutathione s-transferase ("GST") found on a commercial vector. The resulting vector expresses a fusion protein containing an antigenic segment of an *Actinomadura* polyketide synthase and GST that is readily purified from the expressing bacteria using a glutathione affinity column. The purified antigenic fusion protein is used to immunize rabbits. The same approach is used to make antigens based on other segments of *Actinomadura* polyketide synthase. Procedures for making antibodies and for identifying antigenic segments of proteins are well known. See, for instance, Harlow, *Antibodies*, Cold Spring Harbor Press, 1989.

5. Polyketides

In addition to polyketide synthases, the present invention also provides polyketides, including purified pradimicin and pradimicin analogs, and methods for synthesizing polyketides. For example, a
5 vector containing a nucleic acid comprising SEQ ID NO:1 can be expressed in an organism, preferably *Streptomyces*, thereby resulting in pradimicin A synthesis. Preferably, all of the polyketide synthase genes required for polyketide synthesis are present in a single vector, and the genes are preferably in the same configuration as the cDNA.

10 Preferred *Streptomyces* organisms for polyketide synthesis include, for example, *Streptomyces lividans*, *Streptomyces coelicor* and *Streptomyces griseus*. Preferred vectors for expression include, for example, plasmids pIJ61, pIJ702 and pIJ922, which are described in Hopwood et. al., *Gene Manipulation of Streptomyces, A Laboratory*
15 *Manual* (The John Innes Foundation, Norwich, UK 1985). Preferably, the vector includes a promoter that functions well at idiophase, which is a stage of secondary metabolite production, such as the promoter of the *mel* gene, which is present in vector pIJ702.

Preferred methods for preparing a polyketide such as pradimicin or
20 an analog thereof comprise transforming a eukaryotic or prokaryotic cell with an expression vector for expressing intracellularly or extracellularly a nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 70% amino acid identity with an *Actinomadura* polyketide synthase, growing the transformed cell in culture, and isolating the
25 pradimicin or analog thereof from the transformed cell or the culture medium. Preferably, the polypeptide shares at least about 80% amino acid identity with an *Actinomadura* polyketide synthase, and more preferably, the polypeptide shares at least about 90% amino acid identity with an *Actinomadura* polyketide synthase. Most preferably,
30 the expression vector comprises a nucleic acid encoding all polyketide synthase genes necessary for synthesis of pradimicin, such as SEQ ID

NO:1. The production of pradimicin A, for example, can be detected by the presence of a red pigment. Purification of pradimicin from *Actinomadura*, for example, is described in *J. Antibiot.* 41:1701-1704 (1988).

5

The present invention is further exemplified by the following non-limiting example.

Example 1. Cloning of *Actinomadura* Polyketide Synthase Genes

10 **Bacterial strains and plasmids**

Escherichia coli XL1-Blue and pSE101 (*Biosci. Biotech. Biochem.* 59:1835-1841 (1995)), a shuttle cosmid vector replicable in both *Streptomyces lividans* and *E. coli*, were used for preparation of an *Actinomadura hibisca* genomic library. *E. coli* XL1-Blue and plasmids
15 pUC118 and pUC119 were used for sequencing analysis.

DNA isolation and manipulation

Plasmid and genomic DNA isolations were done by the method of Hopwood et. al., *Gene Manipulation of Streptomyces, A Laboratory*
20 *Manual* (The John Innes Foundation, Norwich, UK 1985). Plasmids
-- from *E. coli* were prepared with the Qiagen Plasmid Kit (Qiagen Inc., Chatsworth, CA). All restriction enzymes, T4 ligase and calf intestinal alkaline phosphatase were obtained from Takara (Kyoto, Japan). The procedure for library preparation is described, for example, in *Mol. Gen.*
25 *Genet.* 236:39-48 (1992).

DNA hybridization

The hybridization conditions employed for reactions with the oligonucleotide probe, ³²P-labeled with T4 kinase, were as follows: a
30 Nylon membrane with immobilized DNA was prehybridized at 40°C for 4 hours in 6X SSC buffer, which contains 5X Denhardt's solution

(Maniatis et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1982)), 0.5% SDS and 100 μ g/ml of heat denatured salmon sperm DNA. For overnight hybridization, the same buffer and temperature conditions were used. The genomic DNA

- 5 blotted filter and plasmid DNA blotted filter were washed twice with 6X SSC buffer at 40°C for 30 minutes and with 0.6X SSC buffer at 60°C for 1 hour, respectively.

Cloning of the genes homologous to type II PKS genes

- 10 Amino acid sequences of β -keto synthase, acyl transferase and acyl carrier protein of polyketide synthases are strongly conserved in *Streptomyces* strains producing polyketide antibiotics. See *Annu. Rev. Microbiol.* 47:875-912 (1993) and *J. Biol. Chem.* 267:19278-19290 (1992). Based on these sequences, two oligonucleotide probes were
- 15 synthesized. One was designed based on the amino acid sequences of the *Streptomyces* β -keto synthase around the cysteine residue which is thought to be an active site of the enzyme. See Figure 2, probe 1 (SEQ ID NO:16). The other probe was synthesized based on the amino acid sequences of the *Streptomyces* acyl transferase around the serine
- 20 residue which is believed to be a catalytic domain. See Figure 2, probe 2 (SEQ ID NO:17). Genomic DNA from *Actinomadura hibisca* P157-2 (ATCC 53557) that was digested with several restriction enzymes was subjected to Southern blot analysis with probes 1 and 2, which were separately labeled with 32 P and then mixed. Weak but specific signals
- 25 could be detected. To clone the hybridized fragment, a library was prepared from the strain P157-2 and screened by the colony hybridization with probes 1 and 2 under the same conditions as that for genomic Southern analysis. Several positive cosmid clones were found to hybridize to the probes. Two clones, designated pPRM1 and
- 30 pPRM14, were selected for further analysis.

The physical maps of pPRM1 and pPRM14 were determined and are shown in Figure 3. Using Southern blot hybridization analysis of chromosomal DNA of the strain P-157-2 with these two cosmid clones as probes, it was confirmed that the inserted DNAs of pPRM1 and
5 pPRM14 had not been structurally rearranged during the construction of the library. The position of the hybridized region with oligonucleotide probes was defined by Southern blot analysis.

Sequence analysis.

10 The 8.2-kb *SacI* fragment prepared from pPRM1 was cloned into the *SacI* sites of pUC118 and pUC119 (pUC118 and pUC119 are available, for example, from Takara Syuzo, Kyoto, Japan). After construction of a series of plasmids subcloned from these plasmids, single stranded DNAs were prepared with helper phage M13 KO7,
15 which is also available, for example, from Takara Syuzo. Sequencing was done by the dideoxy chain termination method of Sanger *et al.*, using an automatic DNA sequencer ALF (Pharmacia, Sweden). It was also done with [β -³⁵S]-dCTP as the radioactive label.

20 Nucleotide sequence of the DNA fragment hybridized to the probe

As one approach to examine whether the DNA fragment hybridized to the probes carries the PKS gene for biosynthesis of PRM A, the nucleotide sequence of the 8.2-kb *SacI* fragment containing hybridized region was determined. Computer analysis of the DNA
25 sequence, using Frame Analysis (See *Gene* 30:157-166 (1984)), revealed eleven ORFs (ORF1-11), which are oriented in the same direction except for ORF10. To understand the functions of each the ORFs deduced by DNA sequencing, databases, including DNASIS, were searched using their translated products. The results are summarized in
30 Table 1, *infra*. The ORF1, ORF2 and ORF3 gene products show strong similarities (44-73% amino acid identity) with ORF 1, 2 and 3 gene

products of *gra* (*EMBO J.* 8:2717-2725 (1989)), *tcm* (*EMBO J.* 8:2727-2736 (1989)) and *act* (*J. Biol. Chem.* 267:19278-19290(1992)), which are known to encode condensing enzyme, acyltransferase and acyl carrier protein for granaticin, tetracenomycin and actinorhodin biosynthesis, respectively. The proteins encoded by ORF4 and ORF6 have similarities with the N and C-terminal half of the TcmN protein (*J. Bacteriol.* 174:1810-1820 (1992)) (52% and 46% amino acid identity), respectively, which is thought to be a multifunctional cyclase/dehydratase participating in tetracenomycin biosynthesis. The ORF7 gene product is homologous to the *fabG* product of *E coli* (*J. Biol. Chem.* 267:5751-5754 (1992)) (3-ketoacyl-ACP reductase, 38% amino acid identity) and granaticin-producing polyketide synthase chains 5 and 6 (*EMBO J.* 8:2717-2725 (1989)) (30% and 35% amino acid identity, respectively). Both of the ORF8 and ORF9 gene products have some similarity to hypothetical protein 1 participating in spore color formation in *Streptomyces coelicolor* (*Mol. Microbiol.* 4:1679-1691 (1990)) (23 and 24% amino acid identity, respectively) in a limited region. The ORF10 gene product has a significant similarity to a variety of monooxygenases, including cytochrome P450 (28-40% amino acid identity). The ORF11 gene product shows similarity with the hypothetical protein 1 participating in spore color formation in *Streptomyces coelicolor* (*Mol. Microbiol.* 4:1679-1691 (1990)) (51% amino acid identity), and less extensive, although significant, with the CurG protein of *Streptomyces cyaneus* (*Gene* 117:131-136 (1992)) (45% amino acid identity) and the *tcml* protein of *Streptomyces glaucescens* (EMBL data library no. S27691) (35% amino acid identity). The ORF5 gene product shows some similarity to a histidine kinase of *Caulobacter crescentus* (*Proc. Natl. Acad. Sci.* 89:10297-10301 (1992)) and multicatalytic endopeptidase of *S. cerevisiae* (*Mol. Cell. Biol.* 11:344-353 (1991)).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Oki, Toshikazu
Dairi, Tohru
- (ii) TITLE OF INVENTION: POLYKETIDE SYNTHASES FOR PRADIMICIN
BIOSYNTHESIS AND DNA SEQUENCES ENCODING SAME
- (iii) NUMBER OF SEQUENCES: 25
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Dechert Price & Rhoads
 - (B) STREET: Princeton Pike Corporate Center, PO Box 5218
 - (C) CITY: Princeton
 - (D) STATE: NJ
 - (E) COUNTRY: USA
 - (F) ZIP: 08543-5218
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Bloom, Allen
 - (B) REGISTRATION NUMBER: 29,135
 - (C) REFERENCE/DOCKET NUMBER: BMS-X25
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (609) 520-3214
 - (B) TELEFAX: (609) 520-3259

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8169 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGCTCGGCC	ACGTCGACAC	CGAGGAGCTG	CCCGCCCCCG	ACGAGCAGGG	GCTCGACGTC	60
GGGGGCCGCA	CGTGAGCGGA	CCGAGGGGG	GCGGGCCGCG	CCGCCGTGCG	ATCACCGGCA	120
TGGGGGTGGT	CGCGCCCGGC	GGCTCGGGCC	GGAAGGCGTT	CTGGAACCTG	CTGACCGACG	180
GGCGACCGC	GACCCGGAAG	ATCTCGCTGT	TCGACCCGGC	GGGCTTCCGG	TCCCGGATCG	240
CCGCCGAGTG	CGACTTCGAC	CCCGCCGCCG	AGGGGCTGAC	GCCCCGCGAG	GTCCGGCGCA	300
TGGACCGGGC	CGCGCAGCTC	GCGGTGGTGT	CGGCGCGCGA	GGCGCTCGCC	GACAGCGGGC	360
TGGGGGCGGG	CGAGGGCGAC	CCGGCGCGGT	TCGCGGTGTC	GCTCGGCAGC	GCCGTGCGGT	420
GCACGATGGG	GCTGGAGGAC	GAGTACGTCG	TGGTCAGCGA	CCAGGGCCGC	GACTGGCTGG	480
TCGACCACTC	CTACGGCGTG	CCGCACCTGT	ACCGGCACCT	GGTGCCAGC	TCGCTGGCGG	540
CCGAGGTGCG	CTGGGCGGGC	GGGGCCGAGG	GCCCCGGTCAC	GCTGATCTCG	ACGGGCTCGA	600
CCTCCGGGCT	CGACGCGGTC	GGGCACGGCG	CGCGCGTCAT	CGCCGAGGGC	TCGGCGGACG	660
TGGCGCTCGC	CGGGGCCACC	GACGCGCCCA	TCTCGCCGAT	CACGGTGGCG	TGCTTCGACG	720
CCATCCGGGC	GACCTCGCCG	AACAACGACG	ACCCCGAGCA	CGCGTCCCGG	CCGTTCGACC	780
GGGAGCGCAA	CGGGTTCGTG	CTCGGCGAGG	GCGCGGCGGT	GTTCTGCTCTG	GAGGAGCTGG	840
AGCACGCCCC	CCGCCGGGGC	GCGCACGTCT	ACTGCGAGGT	CGCGGGGTAC	GCCACGCGCG	900
GCAACGCCTA	CCACATGACG	GGCCTGAAGC	CCGACGGCCG	CGAGATGGCC	GAGGCGATCA	960
GGGTGGCGAT	GGACGCCGCC	CGGGTCGCCC	CGGCCGACCT	CGACTACATC	AACGCGCACG	1020
GCTCGGGCAC	CAAGCAGAAC	GACCGGCACG	AGACGGCCGC	GTTCAAGCGC	AGCCTCGGCG	1080
AGCGCGCCTA	CGAGCTGCCG	GTCAGCTCCA	TCAAGTCGAT	GGTCGGGCAC	TCGCTCGGCG	1140
CGATCGGCTC	GATCGAGCTG	GCCGCGTGCG	CGCTGGCGAT	CGAGCACGGT	GTGGTGCCGC	1200
CGACCGCCAA	CCTGCACAAC	GCCGACCCCG	AATGCGACCT	GGACTACGTG	CCGCTGGTGG	1260

CGCGCGAGGG	CCCGATCCGC	ACGGTGCTGA	GCGTGGGCAG	CGGCTTCGGC	GGCTTCCAGT	1320
CCGCCACCGT	CCTGCGGGAG	GCCGCGTGAG	CGTCCTGACG	GCGGACGCGC	CGGCGGTCAC	1380
CGGGATCGGC	GTGGTCGCGC	CGACCGGGAT	CGGCGTCGAG	GAGCACTGGG	CGGCGACGTT	1440
GCGCGGCGTC	CCGGTCATCG	GGCCGCTGAC	CAGGTTTCGAC	GCCGCGCGCT	ACCCGTCGCC	1500
GTTCGGCGGC	GAGGTGCCCC	GGTTCGACGC	CGCCGAGCGC	GTCCCGGGGC	GGCTCATCCC	1560
GCAGACCGAC	CACTGGACGC	ACCTGGCGCT	GGCCGCCACC	GACCTCGCCC	TCGCCGACGC	1620
GGGCGTGCTC	CCGGCCGAGC	TGCCCCAGTA	CGAGATGGCG	GTGGTGACCG	CCAGCTCGTC	1680
GGGCGGCGTG	GAGTTCGGGC	AGCGCGAGAT	CCAGGCGTTG	TGGCGGGACG	GGCCCCGGCA	1740
CGGCGGGGCC	TACCAGTCGA	TCGCCTGGTT	CTACGCGGCG	ACGACCGGCC	AGATCTCCAT	1800
CCGGCACGGG	ATGCGCGGCC	CCTGCGGCGT	CGTGGTCGCC	GAGCAGGCCG	GGGCGCTGGA	1860
GTCTGTCGCG	CAGGCCCCGC	GCTACCTGGC	GGACGGGGCG	CGGGTGGTGG	TGTCCGGCGG	1920
CACCGACGCG	CCGTTCAGTC	CGTACGGCCT	GACCTGCCAG	CTCGGCAGCG	GGCGGCTTAG	1980
CACGGGTGCC	GACCCGGCCC	GCGCCTACCT	GCCGTTTCGAC	GCCGCCGCGA	ACGGCTTCGT	2040
GCCGGGCGAG	GGCGGCGCGA	TCCTCATCAT	CGAGCAAGCC	GCCACCGCGC	AGGACCGCTC	2100
CTACGGGCGG	ATCGCGGGCT	ACGCGGCGAC	CTTCGACCCG	CCGCCGGGCT	CGGGCCGCCC	2160
TCCGACGCTG	GAGCGAGCCG	TGCGCGCCGC	CTTGGACGAC	GCCCGGCTCA	CACCCGCCGA	2220
CGTGGACGTG	GTGTTCGCCC	ACGCGGCGGG	CGTCCCGGAT	CTGGACCGCG	CGGAGGCCGA	2280
CGCGATCGGC	GCGGTCTTCG	GGCCGCGCGG	CGTGCCCGTC	ACCGCGCCCA	AGAGCCTGAC	2340
CGGCCGCCTG	TACGCGGGCG	GCCCCGCGCT	CGACGCGCGG	ACGGCGCTGC	TGGCCATGCA	2400
CGACTCGGTG	ATCCCGCCGA	CGGCCGGCGG	CGCGGACGTC	CCGCCCGGCT	ACGCGCTCGA	2460
CCTGGTCGGC	GCGGAACCGC	GCCCCGCCCG	GCTGCGCACC	GCACTGATCA	TCGCCCCGGG	2520
CTACGGGGGC	TTCAACGCCG	CCCTGGTGCT	GCGCGGCCCC	AACACCTGAC	AACGACCCGA	2580
GAGGACGGAC	GAGATGGCAA	CCCGCGAACG	CACCATCGAC	GACCTGCGCG	CGCTGATGCG	2640
CGCCGCCGTC	GGCGAGGCCG	ACGACATCGA	CCTGGACGGC	GACATCCTCG	ACTCCACCTT	2700
CACCGAGCTG	GAGTACGACT	CGCTCGCCGT	GCTGGAGCTC	GCGGCCCCGA	TCGAGACGCA	2760

GTGGGGCGTG CTGATCCCCG AGGACGACGC GTCCGGGCTG GAGACCCCGC GCATGTTCTT	2820
CGACTACGTG AACGGGCGGG CGGTGGCCGA GCGATGACGC AGTGGCGCAC CGACAGCGTG	2880
ATCGTGATCG ACGCGCCGCT CGACGTCGTC TGGGACATGA CCAACGACGT CGCCTCCTGG	2940
CCGGAGCTGT TCGACGAGTA CGCCTCGGCC GAGATCCTGG AGCGCGACGG CGACACCGTC	3000
CGCTTCCGGC TGACGATGCA CCCCAGCGCC GACGGCAACG CCTGGTCTGT GGTGTCGGAG	3060
CGCACGCCCC ACCGCGCCGC GCTCACCCTC AACCGGCACC GCGTGGAGAC CGGCTGGTTC	3120
GAGCACATGA ACCTGCGCTG GGAATACCGC GAGGTGCCCC GCGGCGTGGA GATGCGCTGG	3180
CGGCAGGACT TCGCGATGAA GGAGGCGTCG CCGGTGTCTG TGGCGGCGAT GACCGAGCGC	3240
ATCCAGAGCA ACTCCCCCGT CCAGATGAAG CTGATCAAGG ACAAGGTGGA GCGGGCGGCC	3300
CGGGGCGCGC GGTGATCGAG TTCTGCTCC CGGTGCGCT GCTCGGCAAC GGGTTGTGCG	3360
CGGGCGTGCT GACGGGCAGC GTCCTCGGCG TCGTGCCGTA CTACCGGACG CTGCCCCGAGG	3420
ACCGCTACAT CGCCGCGCAC GCCTTCGCGG TCGGCCGCTA CGACCCGTTT CAGCCGGTGT	3480
GCCTGCTGGT CACGGTGGCG GCCGACGCGG TCGCGGCGGC GGTGCGGCGG ACCGCCGCGG	3540
CCCGGGTGCT CTGCGCGCTC GCCGCCGTGC TCGCGCTGGC GGTGGTGGCG ATCTCGCTCA	3600
CCCGCAACGT GCCGATGAAC CGCCGGATCA AGCGGCTGGA CCCGGCCGCG CGCCCCGCGG	3660
GGTTCAGCGC GCGCGCGTTC CTGCGCCGCT GGGCGGGCTG GAACGCGGCG CGCACCGGCC	3720
TGACGCTGGC CGCCCTTCTC AGCAACACGG CCGCCCTCGG CGTGCTGCTG TGACCGATCG	3780
GGAAGGGAGG GACATGACCG AACCGGAAGG ACCGCACGCC GCGAGCCTGC GGCTCCAATC	3840
TCTGCTGGAC GGCATGCGCG TCGCCAAGGT CGTCCAGGTG CTCGCCGAAC TCCAGGTGGC	3900
CGACGCGGTC GCCGACGGCC CCTGCAAGCC CGCCGAGATC GCCGCCGACG TCGGCGCCGA	3960
CCCCGACGCG CTGTACCGGG TGCTGCGCTG CGCCGCTCG TTCGGGGTGT TCACCGAGGA	4020
CGAGGACGGC CGGTTCGGGC TCACCCGAT GCCCGCGCTG CTGCGCACCG GCACCGACGA	4080
CAGCCACCGC GACCTGTTCA TGATGGCGGC GGGCGACCTG TGGTGGCGGC CGTACGGCGA	4140
GCTGCTGGAG ACGGTGCGGA CCGGCCGCCC CGCCGCCGAG CTGGCGTTCT GGATGCCGTT	4200
CTACGACTAC CTCGGCACCG ACCCGGCCGC CGCCGGGCTC TTCGACCGCG CGATGACGCA	4260
GGTCAGCAAG GGCCAGGCGA AGGCGATCCT CGGCCGCTGC TCGTTCGAGC GGTACGCGCG	4320

GATCGCCGAC	GTGGGCGGCG	GCCACGGCTA	CTTCCTCGCG	CAGGTGTTGC	GCAGCAGCCC	4380
GCGCACCGAG	GGCGTGCTGC	TGGACCTGCC	GCACGTGGTG	GCCGGAGCCC	CGGCGGTGCT	4440
GGAGAAGCAC	GAGGTCGCCG	ACCGCGTCCA	GGTCGTCCCC	GGCAGCTTCT	TCGACGCGCT	4500
CCCCACGGC	TGCGACGCCT	ACCTGCTGAA	AGCGATCCTC	ATCAACTGGC	CCGACGCCGA	4560
CGCCGAACGC	ATCCTGCACC	GGGTGCCGCA	GGCGATCGGC	AACGACCGCG	ACGCGCGGCT	4620
GCTGGTGGTC	GAGCCCGTCG	TCCCGCCCGG	CGACGTCCGC	GA CTACAGCA	AGGCCACCGA	4680
CATCGACATG	CTCGCCATCA	TCGGCGGGCG	GCAGCGCACC	GTCGCCGAGT	GGCGGCGGCT	4740
GCTGCGCGCG	GGCGGCTTCG	AGCTGCTGGG	CGAGCCACAG	CCGGGCGGCC	GCGAGGTCAT	4800
GGAGTGCCGC	CCCATCTGAA	CCCGTCCCAC	CCGTGCCCCA	CATCCAGGGA	GAACGCATGA	4860
CCGACACATC	GTTCGCCGGC	AAGAACGCGC	TGATCACCGG	CGGCACCCGG	GGCATCGGCC	4920
GGGCGGTGCG	GCTCGGCCTG	GCCGGCGCCG	GGGCCAATGT	CACCGTCTGC	TACCGCAGCG	4980
ACGCCGAGTC	CGCCGCCGCG	ATGGAAGCCG	AGCTGGCCGC	CACCGACGGC	AAGCACCACG	5040
TCCTCCAGGC	CGACATCGGC	AACGCCGGGG	ACGTCCGCCG	CCTGCTGGAC	GAGGTGCGCG	5100
CCCGCATGGG	CTCGCTCGAC	G TAGTCGTGC	ACAACGCCGG	GCTGATCAGC	CACGTGCCGT	5160
TCGCCGACCT	GGAGCCCGAG	GAGTGGCACC	GGATCGTCGA	CTCCAACCTG	ACCGGCATGT	5220
ACCTGGTGGT	GCGGGCCGCG	CTGCCGCTGC	TGTCGGAGGG	CGGCGCGGTC	GTCGGCGTGC	5280
GCTCCAAGGT	CGCGCTCGTC	GGCATCTCGC	AGCGCACCCA	CTACACCGCC	GCCAAGGCCG	5340
GGCTCATCGG	GTTCTGTGCG	TCGCTCAGCA	AGGAGCTGGG	GCCGCTCGGC	ATCCGGGTCA	5400
ACCTGGTTCG	GCCCGGCATC	ACCGAGACCG	ACCAGGCCGC	GCACCTGCCC	CCCGTGCAGC	5460
GCGAGCGCTA	CCAGAGCATG	ACCGCGCTCA	AGCGGCTCGG	CCAGGCCGAC	GAGGTGCGCG	5520
ACGTGGTGCT	GTTCTCTGCC	GGTCCCGGCG	CGCGCTACGT	CACCGGCGAG	ACCGTCAACG	5580
TGGACGGGGG	GATGTGACCA	TGGCCGACAG	CGGCCCGGTG	TTCCGGGTGA	TGCTCCGGAT	5640
GGAGATCGTC	CCGGGCAGGG	AGGCGGAGTT	CGAGCGGCTC	TGGTACTCGG	TCGGCGACAC	5700
CGTCAGCGGC	AACCCCGCCA	ACCTCGGCCA	GTCCGTGCTG	CGCAGCGACG	ACGAGGAGAG	5760
CGTCTACTAC	ATCATGAGCG	ACTGGATCGA	CGAGGCGCGG	TTCCGCGAGT	TCGAGCGCAG	5820

CGACGGCCAC	GTCTAGCACC	GCCGCAAGCT	GCACCCGTAC	CGGGTGAAGG	GCAGCATGGC	5880
GACGATGAAG	GTCGTGCACG	ACCTCGGCCG	CGCGGGCGCG	GAGCCGGTCC	GGTGACGGCC	5940
GGGCAGGTGC	GGGTCCTGGT	CCGCTACCAG	GCTCCGGGCG	ACGACCCCGA	GGCCGTCGTC	6000
CAGGCGTACA	AGCTGGTCTG	CGAGGAACTG	CGCGGGACGC	CCGGCCTGCT	CGGCAGCGAG	6060
CTGCTGGCGT	CGCACGCTCG	ACGAGGGACG	GTTCCGGGTG	CTGAGCCTGT	GGAGCGACGC	6120
CGCGCGGTTT	CAGGAATGGG	AGCAGGGCCC	GGCGCACAAG	GGCCAGACGT	CCGGCCTGCG	6180
CCCGTTCCGG	GACACCTCTT	CGGGGCGCGG	CTTCGATTTC	TACGAAGTGG	TGCACGCCCT	6240
GTAAGAACAA	CGAAGGGCCC	GGCACGCGCA	TGGCGTGCCG	GGCCCTTTCA	CATCCGTGCC	6300
TACCAGGCGA	TGGGCAGCGC	GTCCGGCCGC	GCGAACGCCA	AGCCGGGCGG	CCAGGTGATG	6360
TCGGCATCGT	CGATAGCGAG	ACGCAGCGCG	GGCGTCCGCT	CCACCAGCGT	CTCCAGCACG	6420
ACCTGAAGCT	CCAGCCGGGG	GAGCGGCGCG	CCCAGGCAGT	AGTGGATGCC	GTGGCCGAGC	6480
GCGATGTGCG	GGTTGTCGGT	ACGGCCGAGG	TCGAGTTCCT	CGGGATCGGC	GAACACCTCC	6540
GGATCGCGGT	TGGCGGCGTT	GAAAAGCGGG	ATGACCGCCT	CGCCCGCGCG	CACGAGGGTG	6600
CCGCCGACTT	CCACATCCTC	GACCCGCGATG	CGGATCGCGC	CCGCGCCGCC	GCCGATCTGC	6660
CCGTACCGTA	GCAGTTCCTC	AACGGCCCGC	GGGATACCCG	ACGGGTCCCTC	GCGCAGCCGC	6720
GCGTACCGCG	ACGGCTCGCG	CAGCAGGTGG	TAGACCGAGT	GCGTGATCGC	CGCCGTGGTG	6780
GTGTGGTAAC	CCGCCGCCAG	CAGCGTCATG	CCGAAGGTGA	GCAGTTCCTC	CTCGCTGAGG	6840
CCGTCGTGCG	CGTGCGCCGG	GCTCAGCAAC	GACAGCAGGT	CGTCGGCGGG	CGCGGCCGTC	6900
TTGGCGTCGA	TCAGCTCGGC	GAGGTAGCCG	CGCAGCCGCC	CGACCGCGGC	CTTGATCTCG	6960
TCGGCCTGCG	CGAGAGCGGG	CGCGCCGATG	GTGAGCATCC	GGTCGGTCCA	GTCTTGAAG	7020
CGCGGCCGAT	CCTCCGGCGG	AACGCCCAGC	ATCTCGCAGA	TGACGGTGAC	CGGCAGCGGC	7080
AGCGCCAGGT	GCGCGATCAG	GTCCGCGGGC	GGGCCGTGCT	CGACCATCTC	GTCCACGAAC	7140
CCCGACGTCA	GGTCGCGCAC	GTGCGCGCGC	ATCCCTCCA	CACGACGGGC	GGTGAACGCG	7200
CGAGACACGA	TCTTGCGCAT	CCTCGTGTGC	TCGGGCGGGC	TCATGATGAC	CAGCGACTTG	7260
GAGCCGCGCT	GCATCGGGAT	CAGGCGCGGC	GCGCCCGGCC	GGGTCACCGC	CTCCTTGCTG	7320
AAGCGCCGGT	CCGAGGTGAC	GAACCGGACG	CTGGCGTAGC	GCGTCACGAC	CCACGCGTGG	7380

TCGCCGGTTCG GCAGCACCAC CTTGGCGACC GGGTCGGACG CGCGCAGGCG CGCGTGCTCG	7440
CACGGCGGCT GGAAGGGGTC GTCCGGCCGG AACGGGAAGG CCGGCGTGAC GTCGGGGCGG	7500
GGGTCGACGG TCGGGGCATC CTTTCGAGGAG GGCATACGCC AGGCTTGCAA GGACGCCTCG	7560
AAGCGGGCTC AACCGGGGCT CGCTCCACCG TCCTTCGAGC GGCCCCCGAG CTGCGGTGAC	7620
CACACTCTGC GGCTACCGGC TCACAGCCCC GACCGAGGGA TGTTTCCCAT GGACAGGTTC	7680
CTGATCGTTCG CCCGCATGTC CCCCTCGTCG GAGAAGGAGG TGGCGCGCCT GTTCGCGGAG	7740
TCCGACGAGG GCACCGAGCT GCCGGAGGTG GCCGGGACGG TCAGCCGCAG CCTGCTGTTCG	7800
TTCCACGGCC TGTACTTCCA CCTGACGGAG GTGGAGGAGA GCACGGACAG GACGCTCAAC	7860
GGCATCCACG AACACCCCGA GTTCGTCCGG CTGAGCCGCC AGCTGTCCGG TCACGTCCAG	7920
GCGTACGACC CGAAGACGTG GCGCTCGCCC GCCGACGCCA TGGCCCCGGA GTTCTACCGG	7980
TGGGAGGCGG GGACCGGCGT CGTGCGCCGC TGACCCGTCC CGAGTCCCAC CGGTCCGAGG	8040
TTCTGCTACTC TCCGTTGACT CCCTTCCTCG ATAGCGTCAT CGTTGGTGGC CCACCTGGAC	8100
GACGGAGCCA TCTGAGGGGA AGCGTTGGGT ACCGATACTC TCCCGAGACT CACCGACGCC	8160
GGAGAGCTC	8169

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1278 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGAGCCGAC CGCAGGGGGG CGGGCCGCGC CGCGTCGCGA TCACCGGCAT GGGGGTGGTC	60
GCGCCCGGCG GCTCGGGCCG GAAGGCGTTC TGGAACCTGC TGACCGACGG CCGCACC GCG	120

```

-----
ACCCGGAAGA TCTCGCTGTT CGACCCGGCG GGCTTCCGGT CCCGGATCGC CGCCGAGTGC      180
GACTTCGACC CCGCCGCCGA GGGGCTGACG CCCCgcGAGG TCCGGCGCAT GGACCGGGCC      240
GCGCAGCTCG CGGTGGTGTC GCGCGCGAG GCGCTCGCCG ACAGCGGGCT GGTGGCGGGC      300
GAGGGCGACC CGGCGCGGTT CGCGGTGTCG CTCGGCAGCG CCGTCGGCTG CACGATGGGG      360
CTGGAGGACG AGTACGTCGT GGTcAGCGAC CAGGGCCGCG ACTGGCTGGT CGACCACTCC      420
TACGGCGTGC CGCACCTGTA CCGGCACCTG GTGCCAGCT CGCTGGCGGC CGAGGTCGCC      480
TGGGCGGGCG GGGCCGAGGG CCCGGTCACG CTGATCTCGA CGGGCTGCAC CTCCGGGCTC      540
GACGCGGTcG GGCACGGCGC GCGCGTCATC GCCGAGGGCT CGGCGGACGT GGCGCTCGCC      600
GGGGCCACCG ACcCGCCCAT CTCGCCGATC ACGGTGGCCT GCTTCGACGC CATCCGGGCG      660
ACCTCGCCGA ACAACGACGA CCCCgAGCAC GCGTCCCGGC CGTTCGACCG GGAGCGCAAC      720
GGGTTcGTGC TCGCGGAGGG CGCGGCGGTG TTCGTcCTGG AGGAGCTGGA GCACGCCCCG      780
CGCCGGGGCG CGCACGTCTA CTGCGAGGTC GCGGGGTACG CCACGCGCGG CAACCGCTAC      840
CACATGACGG GCCTGAAGCC CGACGGCCGC GAGATGGCCG AGGCGATCAG GGTGGCGATG      900
GACGCCGCCC GGGTCGCCCC GGCCGACCTC GACTACATCA ACGCGCACGG CTCGGGCACC      960
AAGCAGAACG ACCGGCACGA GACGGCCCGG TTCAAGCGCA GCCTCGGCGA GCGCGCCTAC     1020
GAGCTGCCGG TCAGCTCCAT CAAGTCGATG GTCGGGCACT CGCTCGGCGC GATCGGCTCG     1080
ATCGAGCTGG CCGCGTGCGC GCTGGCGATC GAGCACGGTG TGGTGCCGCC GACCGCCAAC     1140
CTGCACAACG CCGACCCCGA ATGCGACCTG GACTACGTGC CGCTGGTGGC GCGCGAGGGC     1200
CGCATCCGCA CGGTGCTGAG CGTGGGCAGC GGCTTCGGCG GCTTCCAGTC CGCCACCGTC     1260
CTGCGGGAGG CCGCGTGA                                     1278

```

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1223 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTGAGCGTCC TGACGGCGGA CGCGCCGGCG GTCACCGGGA TCGGCGTGGT CGCGCCGACC	60
GGGATCGGCG TCGAGGAGCA CTGGGCGGCG ACGTTGCGCG GCGTCCCGGT CATCGGGCCG	120
CTGACCAGGT TCGACGCCTC GCGCTACCCG TCGCCGTTCTG GCGGCGAGGT GCGCGGGTTC	180
GACGCGCCCG AGCGCGTCCC GGGGCGGCTC ATCCCGCAGA CCGACCACTG GACGCACCTG	240
GCGCTGGCCG CCACCGACCT CGCCCTCGCC GACGCGGGCG TGGTCCCGGC CGAGCTGCCC	300
GAGTACGAGA TGGCGGTGGT GACCGCCAGC TCGTCGGGCG GCGTGGAGTT CGGGCAGCGC	360
GAGATCCAGG CGTTGTGGCG GGACGGGCCC CGGCACGTCG GGGCTACCAG TCGATCGCCT	420
GGTTCTACGC GGCACGACC GGCCAGATCT CCATCCGGCA CGGGATGCGC GCGCCCTGCG	480
GCGTCGTGGT CGCCGAGCAG GCCGGGGCGC TGGAGTCGTT CGCGCAGGCC CGCCGCTACC	540
TGGCGGACGG GCGCGGGTG GTGGTGTCCG GCGGCACCGA CGCGCCGTTT ACTCCGTACG	600
GCCTGACCTG CCAGCTCGGC AGCGGGCGGC TTAGCACGGG TGCCGACCCG GCGCGCGCCT	660
ACCTGCCGTT CGACGCCGCC GCGAACGGCT TCGTGCCGGG CGAGGGCGGC GCGATCCTCA	720
TCATCGAGCA AGCCGCCACC GCGCAGGACC GCTCCTACGG GCGGATCGCG GGCTACGCGG	780
CGACCTTCGA CCCGCCGCCG GGCTCGGGCC GCCCTCCGAC GCTGGAGCGA GCGGTGCGCG	840
CCGCCTTGGA CGACGCCCGG CTCACACCCG CCGACGTGGA CGTGGTGTTT GCGACGCGG	900
CGGGCGTCCC GGATCTGGAC CGCGCGGAGG CCGACGCGAT CGGCGCGGTC TTCGGGCGGC	960
GCGGCGTGCC CGTCACCGCG CCCAAGAGCC TGACCGGCCG CCTGTACGCG GCGGCCCCCG	1020
CGCTCGACGC CGCGACGGCG CTGCTGGCCA TGCACGACTC GGTGATCCCG CCGACGGCCG	1080
GCGGCGCGGA CGTCCCGCCC GGCTACGCGC TCGCCCTGGT CGGCGCGGAA CCGCGCCCCG	1140
CCCGGCTGCG CACCGCACTG ATCATCGCCC GCGGCTACGG GGGCTTCAAC GCGGCCCTGG	1200
TGCTGCGCGG CCCGAACACC TGA	1223

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 264 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGCAACCC GCGAACGCAC CATCGACGAC CTGCGCGCGC TGATGCGCGC CGCCGTCGGC	60
GAGGCCGACG ACATCGACCT GGACGGCGAC ATCCTCGACT CCACCTTCAC CGAGCTGGAG	120
TACGACTCGC TCGCCGTGCT GGAGCTCGCG GCCCGCATCG AGACGCACTG GGGCGTGCTG	180
ATCCCCGAGG ACGACGCGTC CGGGCTGGAG ACCCGCGCA TGTTCCTCGA CTACGTGAAC	240
GGGCGGGCGG TGGCCGAGCG ATGA	264

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATGACGCACT GGCGCACCGA CAGCGTGATC GTGATCGACG CGCCGCTCGA CGTCGTCTGG	60
GACATGACCA ACGACGTCGC CTCCTGGCCG GAGCTGTTCTG ACGAGTACGC CTCGGCCGAG	120

ATCCTGGAGC GCGACGGCGA CACCGTCCGC TTCCGGCTGA CGATGCACCC CGACGCCGAC	180
GGCAACGCCT GGTCTGGGT GTCGGAGCGC ACGCCCGACC GCGCCGCGCT CACCGTCAAC	240
GCGCACCGCG TGGAGACCGG CTGGTTCGAG CACATGAACC TGCCTGGGA CTACCGCGAG	300
GTGCCCCGGCG GCGTGGAGAT GCGCTGGCGG CAGGACTTCG CGATGAAGGA GCGCTCGCCG	360
GTGTCGCTGG CGGCGATGAC CGAGCGCATC CAGAGCAACT CCCCCGTCCA GATGAAGCTG	420
ATCAAGGACA AGGTGGAGCG GGCGGCCCGG GGCGCGCGGT GA	462

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 462 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTGATCGAGT TCCTGCTCCC GGTGCGCTG CTCGGCAACG GGTGTGCGC GGGCGTGCTG	60
ACGGGCAGCG TCCTCGGCGT CGTGCCGTAC TACCGGACGC TGCCCGAGGA CCGCTACATC	120
GCCGCGCAGC CCTTCGCGGT CGGCCGCTAC GACCCGTTCC AGCCGGTGTG CCTGCTGGTC	180
ACGGTGGCGG CCGACGCGGT CGCGGCGGCG GTCGCGCCGA CCGCCGCCGC CCGGGTGCTC	240
TGCGCGCTCG CCGCCGTGCT CGCGCTGGCG GTGGTGGCGA TCTCGCTCAC CCGCAACGTG	300
CCGATGAACC GCCGGATCAA GCGGCTGGAC CCGGCCGCGC CGCCCGCCGG GTTCAGCGCG	360
CCCGCGTTCC TGCGCCGCTG GGCGGGCTGG AACCGGGCGC GCACCGGCCT GACGCTGGCC	420
GCCCTGCTCA GCAACACGGC CGCCCTCGGC GTGCTGCTGT GA	462

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1026 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGACCGAAC CGGAAGGACC GCACGCCGCG AGCCTGCGGC TCCAATCTCT GCTGGACGGC	60
ATGCGCGTCG CCAAGGTCGT GCAGGTGCTC GCCGAACTCC AGGTGGCCGA CGCGGTCGCC	120
GACGGCCCCCT GCAAGCCCCG CGAGATCGCC GCCGACGTCG GCGCCGACCC CGACGCGCTG	180
TACCGGGTGC TGCCTGCGC CGCCTCGTTC GGGGTGTTCA CCGAGGACGA GGACGGCCGG	240
TTCGGGCTCA CCCCAGTGGC CGCGCTGCTG CGCACC GGCA CCGACGACAG CCACCGCGAC	300
CTGTTCATGA TGGCGGCGGG CGACCTGTGG TGGCGGCCGT ACGGCGAGCT GCTGGAGACG	360
GTGCGGACCG GCCGCCCCGC CGCCGAGCTG GCGTTCGGGA TGCCGTTCTA CGACTACCTC	420
GGCACCAGACC CGGCCGCCGC CGGGCTCTTC GACCGCGCGA TGACGCAGGT CAGCAAGGGC	480
CAGGCGAAGG CGATCCTCGG CCGCTGCTCG TTCGAGCGGT ACGCGCGGAT CGCCGACGTG	540
GGCGGCGGCC ACGGCTACTT CCTCGCGCAG GTGTTGCGCA GCAGCCCGCG CACCGAGGGC	600
GTGCTGCTGG ACCTGCCGCA CGTGGTGGCC GGAGCCCCGG CCGTGCTGGA GAAGCACGAG	660
GTGCGCGACC GCGTCCAGGT CGTCCCGGGC AGCTTCTTCG ACGCGCTGCC CACCGGCTGC	720
GACGCCTACC TGCTGAAAGC GATCCTCATC AACTGGCCCG ACGCCGACGC CGAACGCATC	780
CTGCACCGGG TGCGCGAGGC GATCGGCACC GACCGCGACG CGCGGCTGCT GGTGGTCGAG	840
CCCGTCGTCC CGCCCGGCGA CGTCCGCGAC TACAGCAAGG CCACCGACAT CGACATGCTC	900
GCCATCATCG GCGGGCGGCA GCGCACCGTC GCCGAGTGGC GGCGGCTGCT GCGCGCGGGC	960
GGCTTCGAGC TGGTGGGCGA GCCCACGCCG GGCCGCGCG AGGTCATGGA GTGCCGCCCC	1020
ATCTGA	1026

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 741 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGACCGACA CATCGTTCGC CGGCAAGAAC GCGCTGATCA CCGGCGGCAC CCGGGGCATC	60
GGCCGGGGCCG TCGCGCTCGG CCTGGCCCCG GCCGGGGCCA ATGTCACCGT CTGCTACCGC	120
AGCGACGCCG AGTCCGCCGC CGCGATGGAA GCCGAGCTGG CCGCCACCGA CGGCAAGCAC	180
CACGTGCTCC AGGCCGACAT CGGCAACGCC GGGGACGTCC GCCGCCTGCT GGACGAGGTC	240
GGCGCCCCGA TGGGCTCGCT CGACGTAGTC GTGCACAACG CCGGGCTGAT CAGCCACGTG	300
CCGTTCCGGC ACCTGGAGCC CGAGGAGTGG CACCGGATCG TCGACTCAA CCTGACCGGC	360
ATGTACCTGG TGGTGCGGGC CGCGCTGCCG CTGCTGTCTG AGGGCGGCGC GGTCGTCTGGC	420
GTGGGCTCCA AGGTCGCGCT CGTCGGCATC TCGCAGCGCA CCCACTACAC CGCCGCCAAG	480
GCCGGGCTCA TCGGGTTCGT GCGCTCGCTC AGCAAGGAGC TGGGGCCGCT CGGCATCCGG	540
GTCAACCTGG TCGCGCCCGG CATCACCAG ACCGACCAGG CCGCGCACCT GCGCCCCGTG	600
CAGCGCGAGC GCTACCAGAG CATGACCGC CTCAAGCGGC TCGGCCAGGC CGACGAGGTC	660
GCCGACGTGG TGCTGTTTCT CGCCGGTCCC GCGCGCGCT ACGTCACCGG CGAGACCGTC	720
AACGTGGACG GGGGGATGTG A	741

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTGACCATGG CCGACAGCGG CCCGGTGTTC CGGGTGATGC TCCGGATGGA GATCGTCCCG	60
GGCAGGGAGG CGGAGTTCGA GCGGGTCTGG TACTCGGTTC GCGACACCGT CAGCGGCAAC	120
CCCCCAACC TCGGCCAGTG CGTGCTGCCG AGCGACGACG AGGAGAGCGT CTACTACATC	180
ATGAGCGACT GGATCGACGA GCGCGGTTTC CGCGAGTTCG AGCGCAGCGA CGGCCACGTC	240
GAGCACCGCC GCAAGCTGCA CCCGTACCGG GTGAAGGGCA GCATGGCGAC GATGAAGGTC	300
GTGCACGACC TCGGCCGCGC GCGGCGGAG CCGGTCCGGT GA	342

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTGACGGCCG GGCAGGTGCG GGTCTGGTC CGTACCAGG CTCCGGGCGA CGACCCCGAG	60
GCCGTCTGCC AGGCGTACAA GCTGGTCTGC GAGGAACTGC GCGGGACGCC CGGCCTGTCTC	120
GGCAGCGAGC TGCTGGCGTC CACGCTCGAC GAGGGACGGT TCGCGGTGCT GAGCCTGTGG	180

AGCGACGCCG CGCGGTTCCA GGAATGGGAG CAGGGCCCCG CGCACAAGGG CCAGACGTCC	240
GGCCTGCGCC CGTTCCGGGA CACCTCCTCG GGGCGCGGCT TCGATTCTA CGAAGTGGTG	300
CACGCCCTGT AA	312

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1236 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGCCCTCCT CGAAGGATGC CCCGACCGTC GACCCCCGCC CCGACGTCAC GCCGGCCTTC	60
CCGTTCCGGC CGGACGACCC CTTCAGCCG CCGTGCAGC ACGCGGCCT GCGCGCGTCC	120
GACCCGGTCG CCAAGGTGGT GCTGCCGACC GGCGACCACG CGTGGGTCGT GACGCGCTAC	180
GCCGACGTCC GGTTCGTCAC CTCGGACCGG CGCTTCAGCA AGGAGGCGGT GACCCGGCCG	240
GGCGCGCCGC GCCTGATCCC GATGCAGCGC GGCTCCAAGT CGCTGGTCAT CATGGACCCG	300
CCCGAGCACA CGAGGATGCG CAAGATCGTG TCTCGCGCGT TCACCGCCCG TCGTGTGGAG	360
GGGATGCGCG CGCACGTGCG CGACCTGACG TCGGGGTTCTG TGGACGAGAT GGTGAGCAC	420
GGCCCGCCCG CCGACCTGAT CGCGCACCTG GCGCTGCCGC TGCCGGTCAC CGTCATCTGC	480
GAGATGCTGG GCGTTCCGCC GGAGGATCGG CCGCGCTTCC AGGACTGGAC CGACCGGATG	540
CTCACCATCG GCGCGCCCGC TCTCGCGCAG GCCGACGAGA TCAAGGCCGC GGTGCGGCGG	600
CTGCGCGGCT ACCTCGCCGA GCTGATCGAC GCCAAGACGG CCGCGCCCGC CGACGACCTG	660
CTGTGCTTGC TGAGCCGCGC GCACGCCGAC GACGGCCTCA GCGAGGAGGA ACTGCTCACC	720
TTCGGCATGA CGCTGCTGGC GGCGGGTTAC CACACCACCA CGGCGGCGAT CACGCACTCG	780

```

-----
GTCTACCACC TGCTGCGCGA GCCGTCGCGG TACGCGCGGC TCGCGGAGGA CCCGTCGGGT      840
ATCCCGGCGG CCGTTGAGGA ACTGCTACGG TACGGGCAGA TCGGCGGCGG CGCGGGCGCG      900
ATCCGCATCG CGGTCGAGGA TGTGGAAGTC GGCGGCACCC TCGTGC GCGC GGGCGAGGCG      960
GTCATCCCGC TTTTCAACGC CGCCAACCGC GATCCGGAGG TGTTCCGCGA TCCCAGAGAA     1020
CTCGACCTCG GCCGTACCGA CAACCCGCAC ATCGCGCTCG GCCACGGCAT CCACTACTGC      1080
CTGGGCGCGC CGCTCGCCCG GCTGGAGCTT CAGGTCGTGC TGGAGACGCT GGTGGAGCGG      1140
ACGCCCCGCG TCGCTCTCGC TATCGACGAT GCCGACATCA CCTGGCGGCC CGGCTTGCGG      1200
TTCGCGCGGC CGGACGCGCT GCCCATCGCC TGGTAG                                1236

```

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 347 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

ATGGACAGGT TCCTGATCGT CGCCCGCATG TCCCCCTCGT CGGAGAAGGA GGTGGCGCGC      60
CTGTTCCGCG AGTCCGAACG AGGGCACCGA GCTGCCGGAG GTGGCCGGGA CGGTCAGCCG      120
CAGCCTGCTG TCGTTCCACG GCCTGTACTT CCACCTGACG GAGGTGGAGG AGAGCACGGA      180
CAGGACGCTG AACGGCATCC ACGAACACCC CGAGTTCGTC CGGCTGAGCC GCCAGCTGTC      240
CGGTCACGTC CAGGCGTACG AACCCGAAGA CGTGGCGCTC GCGCGCCGAC GCCATGGCCC      300
GCGAGTTCTA CCGGTGGGAG GCGGGGACCG GCGTCGTGCG CCGCTGA                      347

```

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 425 amino acids

(B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met	Ser	Arg	Pro	Gln	Gly	Gly	Gly	Pro	Arg	Arg	Val	Ala	Ile	Thr	Gly	1	5	10	15
Met	Gly	Val	Val	Ala	Pro	Gly	Gly	Ser	Gly	Arg	Lys	Ala	Phe	Trp	Asn	20	25	30	
Leu	Leu	Thr	Asp	Gly	Arg	Thr	Ala	Thr	Arg	Lys	Ile	Ser	Leu	Phe	Asp	35	40	45	
Pro	Ala	Gly	Phe	Arg	Ser	Arg	Ile	Ala	Ala	Glu	Cys	Asp	Phe	Asp	Pro	50	55	60	
Ala	Ala	Glu	Gly	Leu	Thr	Pro	Arg	Glu	Val	Arg	Arg	Met	Asp	Arg	Ala	65	70	75	80
Ala	Gln	Leu	Ala	Val	Val	Ser	Ala	Arg	Glu	Ala	Leu	Ala	Asp	Ser	Gly	85	90	95	
Leu	Val	Ala	Gly	Glu	Gly	Asp	Pro	Ala	Arg	Phe	Ala	Val	Ser	Leu	Gly	100	105	110	
Ser	Ala	Val	Gly	Cys	Thr	Met	Gly	Leu	Glu	Asp	Glu	Tyr	Val	Val	Val	115	120	125	
Ser	Asp	Gln	Gly	Arg	Asp	Trp	Leu	Val	Asp	His	Ser	Tyr	Gly	Val	Pro	130	135	140	
His	Leu	Tyr	Arg	His	Leu	Val	Pro	Ser	Ser	Leu	Ala	Ala	Glu	Val	Ala	145	150	155	160
Trp	Ala	Gly	Gly	Ala	Glu	Gly	Pro	Val	Thr	Leu	Ile	Ser	Thr	Gly	Cys	165	170	175	
Thr	Ser	Gly	Leu	Asp	Ala	Val	Gly	His	Gly	Ala	Arg	Val	Ile	Ala	Glu	180	185	190	

Gly Ser Ala Asp Val Ala Leu Ala Gly Ala Thr Asp Ala Pro Ile Ser
 195 200 205
 Pro Ile Thr Val Ala Cys Phe Asp Ala Ile Arg Ala Thr Ser Pro Asn
 210 215 220
 Asn Asp Asp Pro Glu His Ala Ser Arg Pro Phe Asp Arg Glu Arg Asn
 225 230 235 240
 Gly Phe Val Leu Gly Glu Gly Ala Ala Val Phe Val Leu Glu Glu Leu
 245 250 255
 Glu His Ala Arg Arg Arg Gly Ala His Val Tyr Cys Glu Val Ala Gly
 260 265 270
 Tyr Ala Thr Arg Gly Asn Ala Tyr His Met Thr Gly Leu Lys Pro Asp
 275 280 285
 Gly Arg Glu Met Ala Glu Ala Ile Arg Val Ala Met Asp Ala Ala Arg
 290 295 300
 Val Ala Pro Ala Asp Leu Asp Tyr Ile Asn Ala His Gly Ser Gly Thr
 305 310 315 320
 Lys Gln Asn Asp Arg His Glu Thr Ala Ala Phe Lys Arg Ser Leu Gly
 325 330 335
 Glu Arg Ala Tyr Glu Leu Pro Val Ser Ser Ile Lys Ser Met Val Gly
 340 345 350
 His Ser Leu Gly Ala Ile Gly Ser Ile Glu Leu Ala Ala Cys Ala Leu
 355 360 365
 Ala Ile Glu His Gly Val Val Pro Pro Thr Ala Asn Leu His Asn Ala
 370 375 380
 Asp Pro Glu Cys Asp Leu Asp Tyr Val Pro Leu Val Ala Arg Glu Gly
 385 390 395 400
 Arg Ile Arg Thr Val Leu Ser Val Gly Ser Gly Phe Gly Gly Phe Gln
 405 410 415
 Ser Ala Thr Val Leu Arg Glu Ala Ala
 420 425

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met	Ser	Val	Leu	Thr	Ala	Asp	Ala	Pro	Ala	Val	Thr	Gly	Ile	Gly	Val	1	5	10	15
Val	Ala	Pro	Thr	Gly	Ile	Gly	Val	Glu	Glu	His	Trp	Ala	Ala	Thr	Leu	20	25	30	
Arg	Gly	Val	Pro	Val	Ile	Gly	Pro	Leu	Thr	Arg	Phe	Asp	Ala	Ser	Arg	35	40	45	
Tyr	Pro	Ser	Pro	Phe	Gly	Gly	Glu	Val	Pro	Gly	Phe	Asp	Ala	Ala	Glu	50	55	60	
Arg	Val	Pro	Gly	Arg	Leu	Ile	Pro	Gln	Thr	Asp	His	Trp	Thr	His	Leu	65	70	75	80
Ala	Leu	Ala	Ala	Thr	Asp	Leu	Ala	Leu	Ala	Asp	Ala	Gly	Val	Val	Pro	85	90	95	
Ala	Glu	Leu	Pro	Glu	Tyr	Glu	Met	Ala	Val	Val	Thr	Ala	Ser	Ser	Ser	100	105	110	
Gly	Gly	Val	Glu	Phe	Gly	Gln	Arg	Glu	Ile	Gln	Ala	Leu	Trp	Arg	Asp	115	120	125	
Gly	Pro	Arg	His	Val	Gly	Ala	Tyr	Gln	Ser	Ile	Ala	Trp	Phe	Tyr	Ala	130	135	140	
Ala	Thr	Thr	Gly	Gln	Ile	Ser	Ile	Arg	His	Gly	Met	Arg	Gly	Pro	Cys	145	150	155	160
Gly	Val	Val	Val	Ala	Glu	Gln	Ala	Gly	Ala	Leu	Glu	Ser	Phe	Ala	Gln	165	170	175	
Ala	Arg	Arg	Tyr	Leu	Ala	Asp	Gly	Ala	Arg	Val	Val	Val	Ser	Gly	Gly	180	185	190	
Thr	Asp	Ala	Pro	Phe	Ser	Pro	Tyr	Gly	Leu	Thr	Cys	Gln	Leu	Gly	Ser	195	200	205	

Gly Arg Leu Ser Thr Gly Ala Asp Pro Ala Arg Ala Tyr Leu Pro Phe
 210 215 220
 Asp Ala Ala Ala Asn Gly Phe Val Pro Gly Glu Gly Gly Ala Ile Leu
 225 230 235 240
 Ile Ile Glu Gln Ala Ala Thr Ala Gln Asp Arg Ser Tyr Gly Arg Ile
 245 250 255
 Ala Gly Tyr Ala Ala Thr Phe Asp Pro Pro Pro Gly Ser Gly Arg Pro
 260 265 270
 Pro Thr Leu Glu Arg Ala Val Arg Ala Ala Leu Asp Asp Ala Arg Leu
 275 280 285
 Thr Pro Ala Asp Val Asp Val Val Phe Ala Asp Ala Ala Gly Val Pro
 290 295 300
 Asp Leu Asp Arg Ala Glu Ala Asp Ala Ile Gly Ala Val Phe Gly Pro
 305 310 315 320
 Arg Gly Val Pro Val Thr Ala Pro Lys Ser Leu Thr Gly Arg Leu Tyr
 325 330 335
 Ala Gly Gly Pro Ala Leu Asp Ala Ala Thr Ala Leu Leu Ala Met His
 340 345 350
 Asp Ser Val Ile Pro Pro Thr Ala Gly Gly Ala Asp Val Pro Pro Gly
 355 360 365
 Tyr Ala Leu Asp Leu Val Gly Ala Glu Pro Arg Pro Ala Arg Leu Arg
 370 375 380
 Thr Ala Leu Ile Ile Ala Arg Gly Tyr Gly Gly Phe Asn Ala Ala Leu
 385 390 395 400
 Val Leu Arg Gly Pro Asn Thr
 405

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

Met Ala Thr Arg Glu Arg Thr Ile Asp Asp Leu Arg Ala Leu Met Arg
1          5          10          15
Ala Ala Val Gly Glu Ala Asp Asp Ile Asp Leu Asp Gly Asp Ile Leu
20          25          30
Asp Ser Thr Phe Thr Glu Leu Glu Tyr Asp Ser Leu Ala Val Leu Glu
35          40          45
Leu Ala Ala Arg Ile Glu Thr Gln Trp Gly Val Leu Ile Pro Glu Asp
50          55          60
Asp Ala Ser Gly Leu Glu Thr Pro Arg Met Phe Leu Asp Tyr Val Asn
65          70          75          80
Gly Arg Ala Val Ala Glu Arg
85

```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 153 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```

Met Thr Gln Trp Arg Thr Asp Ser Val Ile Val Ile Asp Ala Pro Leu
1          5          10          15
Asp Val Val Trp Asp Met Thr Asn Asp Val Ala Ser Trp Pro Glu Leu
20          25          30

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 153 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(iv) ANTI-SENSE: NO

Met Ile Glu Phe Leu Leu Pro Val Ala Leu Leu Gly Asn Gly Leu Cys
1 5 10 15
Ala Gly Val Leu Thr Gly Ser Val Leu Gly Val Val Pro Tyr Tyr Arg
20 25 30
Thr Leu Pro Glu Asp Arg Tyr Ile Ala Ala His Ala Phe Ala Val Gly
35 40 45

Arg Tyr Asp Pro Phe Gln Pro Val Cys Leu Leu Val Thr Val Ala Ala
 50 55 60
 Asp Ala Val Ala Ala Ala Val Ala Pro Thr Ala Ala Ala Arg Val Leu
 65 70 75 80
 Cys Ala Leu Ala Ala Val Leu Ala Leu Ala Val Val Ala Ile Ser Leu
 85 90 95
 Thr Arg Asn Val Pro Met Asn Arg Arg Ile Lys Arg Leu Asp Pro Ala
 100 105 110
 Ala Pro Pro Ala Gly Phe Ser Ala Pro Ala Phe Leu Arg Arg Trp Ala
 115 120 125
 Gly Trp Asn Ala Ala Arg Thr Gly Leu Thr Leu Ala Ala Leu Leu Ser
 130 135 140
 Asn Thr Ala Ala Leu Gly Val Leu Leu
 145 150

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 341 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Thr Glu Pro Glu Gly Pro His Ala Ala Ser Leu Arg Leu Gln Ser
 1 5 10 15
 Leu Leu Asp Gly Met Arg Val Ala Lys Val Val Gln Val Leu Ala Glu
 20 25 30
 Leu Gln Val Ala Asp Ala Val Ala Asp Gly Pro Cys Lys Pro Ala Glu
 35 40 45
 Ile Ala Ala Asp Val Gly Ala Asp Pro Asp Ala Leu Tyr Arg Val Leu

50		55		60													
Arg 65	Cys	Ala	Ala	Ser	Phe 70	Gly	Val	Phe	Thr	Glu 75	Asp	Glu	Asp	Gly	Arg 80		
Phe	Gly	Leu	Thr	Pro 85	Met	Ala	Ala	Leu	Leu 90	Arg	Thr	Gly	Thr	Asp 95	Asp		
Ser	His	Arg	Asp 100	Leu	Phe	Met	Met	Ala 105	Ala	Gly	Asp	Leu	Trp 110	Trp	Arg		
Pro	Tyr	Gly 115	Glu	Leu	Leu	Glu	Thr	Val	Arg	Thr	Gly	Arg 125	Pro	Ala	Ala		
Glu 130	Leu	Ala	Phe	Gly	Met	Pro 135	Phe	Tyr	Asp	Tyr	Leu 140	Gly	Thr	Asp	Pro		
Ala 145	Ala	Ala	Gly	Leu	Phe 150	Asp	Arg	Ala	Met	Thr 155	Gln	Val	Ser	Lys	Gly 160		
Gln	Ala	Lys	Ala	Ile 165	Leu	Gly	Arg	Cys	Ser 170	Phe	Glu	Arg	Tyr	Ala 175	Arg		
Ile	Ala	Asp	Val 180	Gly	Gly	Gly	His	Gly 185	Tyr	Phe	Leu	Ala	Gln 190	Val	Leu		
Arg	Ser	Ser 195	Pro	Arg	Thr	Glu	Gly 200	Val	Leu	Leu	Asp	Leu 205	Pro	His	Val		
Val 210	Ala	Gly	Ala	Pro	Ala	Val 215	Leu	Glu	Lys	His	Glu 220	Val	Ala	Asp	Arg		
Val 225	Gln	Val	Val	Pro	Gly 230	Ser	Phe	Phe	Asp	Ala 235	Leu	Pro	Thr	Gly	Cys 240		
Asp	Ala	Tyr	Leu	Leu 245	Lys	Ala	Ile	Leu	Ile 250	Asn	Trp	Pro	Asp	Ala 255	Asp		
Ala	Glu	Arg	Ile 260	Leu	His	Arg	Val	Arg	Glu 265	Ala	Ile	Gly	Thr	Asp 270	Arg		
Asp	Ala	Arg 275	Leu	Leu	Val	Val	Glu 280	Pro	Val	Val	Pro	Pro 285	Gly	Asp	Val		
Arg 290	Asp	Tyr	Ser	Lys	Ala	Thr 295	Asp	Ile	Asp	Met	Leu 300	Ala	Ile	Ile	Gly		
Gly 305	Arg	Gln	Arg	Thr	Val 310	Ala	Glu	Trp	Arg	Arg 315	Leu	Leu	Arg	Ala	Gly 320		
Gly	Phe	Glu	Leu	Val	Gly	Glu	Pro	Thr	Pro	Gly	Arg	Arg	Glu	Val	Met		

325	330	335
Glu Cys Arg Pro Ile		
340		

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Thr	Asp	Thr	Ser	Phe	Ala	Gly	Lys	Asn	Ala	Leu	Ile	Thr	Gly	Gly	1	5	10	15
Thr	Arg	Gly	Ile	Gly	Arg	Ala	Val	Ala	Leu	Gly	Leu	Ala	Arg	Ala	Gly	20	25	30	
Ala	Asn	Val	Thr	Val	Cys	Tyr	Arg	Ser	Asp	Ala	Glu	Ser	Ala	Ala	Ala	35	40	45	
Met	Glu	Ala	Glu	Leu	Ala	Ala	Thr	Asp	Gly	Lys	His	His	Val	Leu	Gln	50	55	60	
Ala	Asp	Ile	Gly	Asn	Ala	Gly	Asp	Val	Arg	Arg	Leu	Leu	Asp	Glu	Val	65	70	75	80
Ala	Ala	Arg	Met	Gly	Ser	Leu	Asp	Val	Val	Val	His	Asn	Ala	Gly	Leu	85	90	95	
Ile	Ser	His	Val	Pro	Phe	Ala	Asp	Leu	Glu	Pro	Glu	Glu	Trp	His	Arg	100	105	110	
Ile	Val	Asp	Ser	Asn	Leu	Thr	Gly	Met	Tyr	Leu	Val	Val	Arg	Ala	Ala	115	120	125	
Leu	Pro	Leu	Leu	Ser	Glu	Gly	Gly	Ala	Val	Val	Gly	Val	Gly	Ser	Lys	130	135	140	

Val Ala Leu Val Gly Ile Ser Gln Arg Thr His Tyr Thr Ala Ala Lys
 145 150 155 160
 Ala Gly Leu Ile Gly Phe Val Arg Ser Leu Ser Lys Glu Leu Gly Pro
 165 170 175
 Leu Gly Ile Arg Val Asn Leu Val Ala Pro Gly Ile Thr Glu Thr Asp
 180 185 190
 Gln Ala Ala His Leu Pro Pro Val Gln Arg Glu Arg Tyr Gln Ser Met
 195 200 205
 Thr Ala Leu Lys Arg Leu Gly Gln Ala Asp Glu Val Ala Asp Val Val
 210 215 220
 Leu Phe Leu Ala Gly Pro Gly Ala Arg Tyr Val Thr Gly Glu Thr Val
 225 230 235 240
 Asn Val Asp Gly Gly Met
 245

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 113 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Val Thr Met Ala Asp Ser Gly Pro Val Phe Arg Val Met Leu Arg Met
 1 5 10 15
 Glu Ile Val Pro Gly Arg Glu Ala Glu Phe Glu Arg Val Trp Tyr Ser
 20 25 30
 Val Gly Asp Thr Val Ser Gly Asn Pro Ala Asn Leu Gly Gln Cys Val
 35 40 45
 Leu Arg Ser Asp Asp Glu Glu Ser Val Tyr Tyr Ile Met Ser Asp Trp
 50 55 60

Ile Asp Glu Ala Arg Phe Arg Glu Phe Glu Arg Ser Asp Gly His Val
 65 70 75 80
 Glu His Arg Arg Lys Leu His Pro Tyr Arg Val Lys Gly Ser Met Ala
 85 90 95
 Thr Met Lys Val Val His Asp Leu Gly Arg Ala Ala Ala Glu Pro Val
 100 105 110
 Arg

(2) INFORMATION FOR SEQ ID NO:21:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 103 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Val Thr Ala Gly Gln Val Arg Val Leu Val Arg Tyr Gln Ala Pro Gly
 1 5 10 15
 Asp Asp Pro Glu Ala Val Val Gln Ala Tyr Lys Leu Val Cys Glu Glu
 20 25 30
 Leu Arg Gly Thr Pro Gly Leu Leu Gly Ser Glu Leu Leu Ala Ser Thr
 35 40 45
 Leu Asp Glu Gly Arg Phe Ala Val Leu Ser Leu Trp Ser Asp Ala Ala
 50 55 60
 Arg Phe Gln Glu Trp Glu Gln Gly Pro Ala His Lys Gly Gln Thr Ser
 65 70 75 80
 Gly Leu Arg Pro Phe Arg Asp Thr Ser Ser Gly Arg Gly Phe Asp Phe
 85 90 95
 Tyr Glu Val Val His Ala Leu

100

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 411 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```

Met Pro Ser Ser Lys Asp Ala Pro Thr Val Asp Pro Arg Pro Asp Val
1          5          10          15
Thr Pro Ala Phe Pro Phe Arg Pro Asp Asp Pro Phe Gln Pro Pro Cys
          20          25          30
Glu His Ala Arg Leu Arg Ala Ser Asp Pro Val Ala Lys Val Val Leu
          35          40          45
Pro Thr Gly Asp His Ala Trp Val Val Thr Arg Tyr Ala Asp Val Arg
          50          55          60
Phe Val Thr Ser Asp Arg Arg Phe Ser Lys Glu Ala Val Thr Arg Pro
65          70          75          80
Gly Ala Pro Arg Leu Ile Pro Met Gln Arg Gly Ser Lys Ser Leu Val
          85          90          95
Ile Met Asp Pro Pro Glu His Thr Arg Met Arg Lys Ile Val Ser Arg
          100          105          110
Ala Phe Thr Ala Arg Arg Val Glu Gly Met Arg Ala His Val Arg Asp
          115          120          125
Leu Thr Ser Gly Phe Val Asp Glu Met Val Glu His Gly Pro Pro Ala
          130          135          140
Asp Leu Ile Ala His Leu Ala Leu Pro Leu Pro Val Thr Val Ile Cys
145          150          155          160

```

Glu Met Leu Gly Val Pro Pro Glu Asp Arg Pro Arg Phe Gln Asp Trp
 165 170 175
 Thr Asp Arg Met Leu Thr Ile Gly Ala Pro Ala Leu Ala Gln Ala Asp
 180 185 190
 Glu Ile Lys Ala Ala Val Gly Arg Leu Arg Gly Tyr Leu Ala Glu Leu
 195 200 205
 Ile Asp Ala Lys Thr Ala Ala Pro Ala Asp Asp Leu Leu Ser Leu Leu
 210 215 220
 Ser Arg Ala His Ala Asp Asp Gly Leu Ser Glu Glu Glu Leu Leu Thr
 225 230 235 240
 Phe Gly Met Thr Leu Leu Ala Ala Gly Tyr His Thr Thr Thr Ala Ala
 245 250 255
 Ile Thr His Ser Val Tyr His Leu Leu Arg Glu Pro Ser Arg Tyr Ala
 260 265 270
 Arg Leu Arg Glu Asp Pro Ser Gly Ile Pro Ala Ala Val Glu Glu Leu
 275 280 285
 Leu Arg Tyr Gly Gln Ile Gly Gly Gly Ala Gly Ala Ile Arg Ile Ala
 290 295 300
 Val Glu Asp Val Glu Val Gly Gly Thr Leu Val Arg Ala Gly Glu Ala
 305 310 315 320
 Val Ile Pro Leu Phe Asn Ala Ala Asn Arg Asp Pro Glu Val Phe Ala
 325 330 335
 Asp Pro Glu Glu Leu Asp Leu Gly Arg Thr Asp Asn Pro His Ile Ala
 340 345 350
 Leu Gly His Gly Ile His Tyr Cys Leu Gly Ala Pro Leu Ala Arg Leu
 355 360 365
 Glu Leu Gln Val Val Leu Glu Thr Leu Val Glu Arg Thr Pro Ala Leu
 370 375 380
 Arg Leu Ala Ile Asp Asp Ala Asp Ile Thr Trp Arg Pro Gly Leu Ala
 385 390 395 400
 Phe Ala Arg Pro Asp Ala Leu Pro Ile Ala Trp
 405 410

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

-
- (A) LENGTH: 114 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

Met Asp Arg Phe Leu Ile Val Ala Arg Met Ser Pro Ser Ser Glu Lys
1           5           10           15
Glu Val Ala Arg Leu Phe Ala Glu Ser Asp Glu Gly Thr Glu Leu Pro
          20           25           30
Glu Val Ala Gly Thr Val Ser Arg Ser Leu Leu Ser Phe His Gly Leu
          35           40           45
Tyr Phe His Leu Thr Glu Val Glu Glu Ser Thr Asp Arg Thr Leu Asn
          50           55           60
Gly Ile His Glu His Pro Glu Phe Val Arg Leu Ser Arg Gln Leu Ser
          65           70           75           80
Gly His Val Gln Ala Tyr Asp Pro Lys Thr Trp Arg Ser Pro Ala Asp
          85           90           95
Ala Met Ala Arg Glu Phe Tyr Arg Trp Glu Ala Gly Thr Gly Val Val
          100          105          110
Arg Arg

```

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "probe"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGCGCGGAGG GCGCGGTCAC GATGGTCTCC ACCGGCTGCA CCTCGGGCCT GGAC

54

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "probe"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCCGTCAGCT CCATCAAGTC CATGGTCGGC CACTCGCTCG GCGCGATCGG CTCC

54

WE CLAIM:

1. A substantially pure nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 75% amino acid identity with an *Actinomadura* polyketide synthase.

5

2. The nucleic acid of claim 1, encoding a polypeptide sharing at least about 80% amino acid identity with an *Actinomadura* polyketide synthase.

10

3. The nucleic acid of claim 2, encoding a polypeptide sharing at least about 90% amino acid identity with an *Actinomadura* polyketide synthase.

15

4. The substantially pure nucleic acid of claim 1, comprising a nucleic acid selected from the group consisting of SEQ ID NO:1-12.

5. A transformed eukaryotic or prokaryotic cell comprising the nucleic acid of claim 1.

20

6. A vector capable of reproducing in a eukaryotic or prokaryotic cell comprising the nucleic acid of claim 1.

7. A substantially pure nucleic acid comprising a nucleic acid that hybridizes to the nucleic acid of claim 1 under stringent conditions.

25

8. A substantially pure nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 75% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone.

30

9. The substantially pure nucleic acid of claim 8, encoding a polypeptide sharing at least about 80% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone.

35 10. The nucleic acid of claim 9, encoding a polypeptide sharing at least about 90% amino acid identity with a polyketide synthase for biosynthesis of a benzo(a)naphthacenequinone.

40 11. The nucleic acid of claim 10, wherein the polyketide synthase is an *Actinomadura* polyketide synthase.

12. The nucleic acid of claim 11, wherein the polyketide synthase is an *Actinomadura* polyketide synthase.

45 13. The nucleic acid of claim 12, wherein the polyketide synthase is an *Actinomadura* polyketide synthase.

50 14. The nucleic acid of claim 8, wherein the benzo(a)naphthacenequinone is a dihydrobenzo(a)naphthacenequinone aglycon.

55 15. The nucleic acid of claim 9, wherein the benzo(a)naphthacenequinone is a dihydrobenzo(a)naphthacenequinone aglycon.

16. The nucleic acid of claim 10, wherein the benzo(a)naphthacenequinone is a dihydrobenzo(a)naphthacenequinone aglycon.

60 17. The nucleic acid of claim 14, wherein the dihydrobenzo(a)naphthacenequinone aglycon is pradimicin.

18. The nucleic acid of claim 15, wherein the dihydrobenzo(a)naphthacenequinone aglycon is pradimicin.

65 19. The nucleic acid of claim 16, wherein the dihydrobenzo(a)naphthacenequinone aglycon is pradimicin.

20. A substantially pure polypeptide comprising an amino acid sequence sharing at least about 75% amino acid identity with an
70 *Actinomadura* polyketide synthase.

21. The polypeptide of claim 20, comprising an amino acid sequence sharing at least about 80% amino acid identity with an *Actinomadura* polyketide synthase.

22. The polypeptide of claim 21, comprising an amino acid sequence sharing at least about 90% amino acid identity with an *Actinomadura* polyketide synthase.

23. The polypeptide of claim 22, comprising an amino acid sequence selected from the group consisting of SEQ ID NO:13, SEQ ID NO:14 and SEQ ID NO:15.

24. A method of preparing pradimicin or an analog thereof comprising:

(a) transforming a eukaryotic or prokaryotic cell with an expression vector for expressing intracellularly or extracellularly a nucleic acid comprising a nucleic acid encoding a polypeptide sharing at least about 70% amino acid identity with an *Actinomadura* polyketide synthase;

(b) growing the transformed cell in culture; and

(c) isolating the pradimicin or analog thereof from the transformed cell or the culture medium.

25. The method of claim 24, wherein the polypeptide shares at least about 80% amino acid identity with an *Actinomadura* polyketide synthase.

26. The method of claim 25, wherein the polypeptide shares at least about 90% amino acid identity with an *Actinomadura* polyketide synthase.

27. The method of claim 24, wherein the nucleic acid comprises SEQ ID NO:1.

1/5

Pradimicin A; R1=H, R2=H
Pradimicin S; R1=CH₂OH, R2=HO₃S

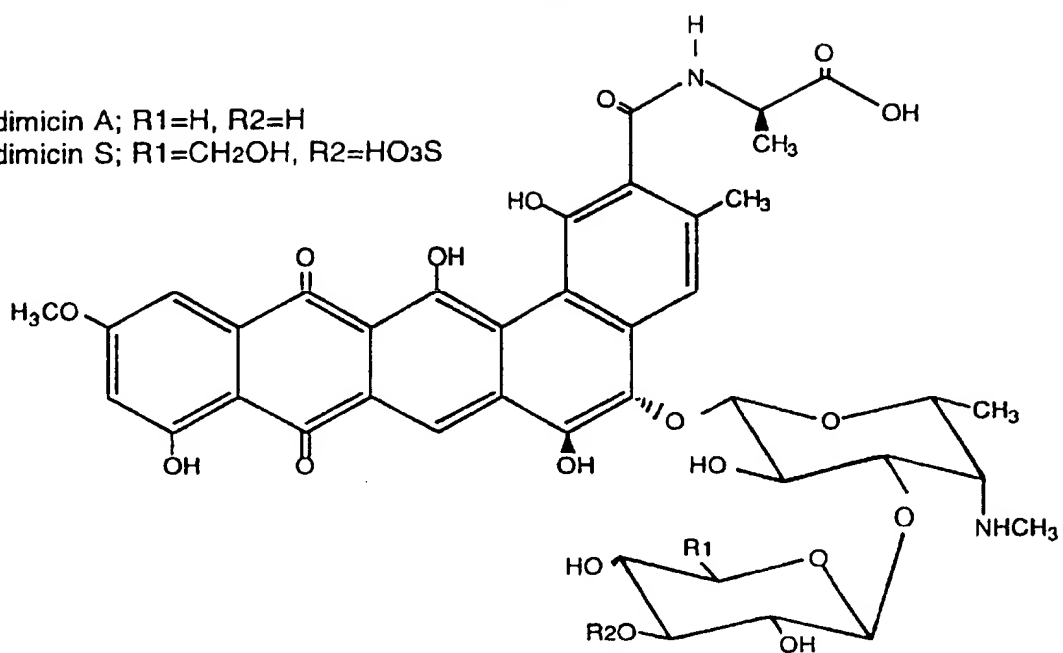


FIGURE 1

β - Keto synthase

Granatidin	G	A	E	G	P	V	T	M	V	S	D	G	C	T	S	G	L	D
Tetracenomycin	G	A	E	G	P	V	T	V	V	S	T	G	C	T	S	G	L	D
Actinorhodin	G	A	E	G	P	V	T	M	V	S	T	G	C	T	S	G	L	D
CONSENSUS	G	A	E	G	P	V	T	M	V	S	T	G	C	T	S	G	L	D

Probe 1 (54 mer) 5' -GGCGCGGAGGGCCCGGTCACGATGGTCTCCACCGGCTGCACCTCGGGCCTGGAC-3'

Acyl transferase

Granatidin	P	V	S	S	I	K	S	M	G	G	H	S	L	G	A	I	G	S
Tetracenomycin	P	V	S	S	I	K	S	M	I	G	H	S	L	G	A	I	G	S
Actinorhodin	P	V	S	S	I	K	S	M	V	G	H	S	L	G	A	I	G	S
CONSENSUS	P	V	S	S	I	K	S	M	()	G	H	S	L	G	A	I	G	S

Probe 2 (54 mer) 5' -CCCGTCAGCTCCATCAAGTCCATGGTCGGCCACTCGCTCGGCGCGATCGGCTCC-3'

FIGURE 2

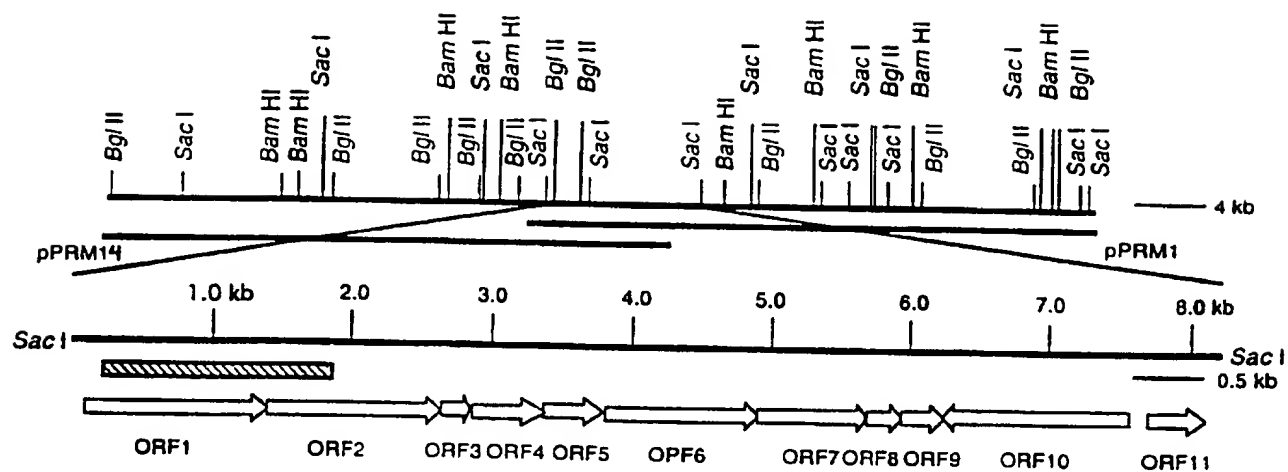


FIGURE 3

1 60
 A MSRPQGGGPRRVAITGMGVVAPGGSGRKAFWNLLTDGRTATRKISLFDPAGFRSRIAAEC
 . ** . ** . ** . ** . * . ** . ** . ** . ** . * . ** . ** . *
 B MTRHAEKRVIITGIGVRAPGGAGTAAFWDLLTAGRTATRTISLFDAAPYRSRIAGEI
 1 57
 DFDPAAEGLTPREVRMDRAAQLAVVSAREALADSGLVAGEGDPARFAVSLGSVAGCTMG
 * . ** . ** . * . ** . ** . ** . ** . * . ** . ** . * . ** . * . ** . *
 DFDPIGEGLSPROASTYDRATQLAVVCAREALKDSGLDPAAVNPERIGVSI GTAVGCTTG

 LEDEYVVVSDQGRDWLVHSGYGVPHLYRHLVPSSLAAEVAWAGGAEGPVTLSIGCTSGL
 * . ** . ** . * . ** . ** . * . ** . * . ** . ** . ** . ** . ** .
 LDREYARVSEGGSRWLVDHTLAVEQLFDYFVPTSICREVAWEAGAEGPVTVVSTGCTSGL

 DAVGHGARVIAEGSADVALAGATDAPISPI TVACFDAIRATSPNDDPEHASRPFDRERN
 * . ** . * . * . * . ** . ** . ** . ** . ** . ** . ** . ** . * .
 DAVGYGTELIRDGRADVVCATDAPISPI TVACFDAIKATSANNDPAHASRPFDRNRD

 GFVLGEGAAVFVLEELHARRRGHAHVYCEVAGYATRGNAYHMTGLKPDGREMAEIRVAM
 * . ** . ** . ** . ** . ** . * . ** . * . ** . * . ** . ** . ** . * .
 GFVLGEGSAVFVLEELSAARRRGHAHAYAEVRGFATRSNAFHMTGLKPDGREMAEITAAL

 DAARVAPADLDYINAHGSGTKQNDRHETAAFKRSLGERAYELPVSSIKSMVGHSLGAIGS
 * . ** . * . ** . ** . ** . ** . ** . ** . ** . ** . ** . ** .
 DQARRTGDDLHYINAHGSGTRQNDRHETAAFKRSLGQRAYDVPVSSIKSMIGHSLGAIGS

 IELAACALAEHGVVPPTANLHNADPECDLDYVPLVAREGRIRTVLSVGSFGGFGQSATV
 * . ** . ** . ** . ** . ** . * . ** . * . ** . * . ** . ** . ** . * .
 IELAACALAEHGVIPPTANYEEDPECDLDYVPNVAREQVRDVTLSVGSFGGFGQSAAV
 425
 LREAA
 *
 LARPK
 422

FIGURE 4

FIGURE 5

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 96/14791

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/52 C12N1/21 C12N9/00 C12P21/00 //(C12N1/21, C12R1:03)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MOL. GEN. GENET., vol. 251, 1996, pages 113-120, XP000652375 K. YLIHONKO ET AL.: "A gene cluster involved in nogalamycin biosynthesis from Streptomyces nogalater: sequence analysis and complementation of early-block mutations in the anthracycline pathway" see the whole document.</p> <p style="text-align: center;">--- -/--</p>	<p>1,6-8, 20,24</p>

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 July 1997	04.08.97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Authorized officer Yeats, S

INTERNATIONAL SEARCH REPORT

Inter. nal Application No

PCT/US 96/14791

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MOL. GEN. GENET., vol. 240, 1993, pages 146-150, XP000654921 C. LE GOUILL ET AL.: "Saccharopolyspora hirsuta 367 encodes clustered genes similar to ketoacyl synthase, ketoacyl reductase, acyl carrier protein, and biotin carboxyl carrier protein" see the whole document. ---</p>	1,6-8, 20,24
Y	<p>J. BIOL. CHEM., vol. 267, 1992, pages 19278-19290, XP000652285 M.A. FERNANDEZ-MORENO ET AL.: "Nucleotide sequence and deduced functions of a set of cotranscribed genes of Streptomyces coelicolor A3(2) including the polyketide synthase for the antibiotic actinorhodin" cited in the application see the whole document, especially the abstract and Figure 4. ---</p>	1-27
Y	<p>J. ANTIBIOTICS, vol. 48, 1995, pages 162-168, XP000654920 K. SAITOH ET AL.: "Pradimicin S, a new pradimicin analog. III. Application of the Frit-FAB LC/MS technique to the elucidation of the pradimicin S biosynthetic pathway" see the whole document. -----</p>	1-27